MOLECULAR MECHANISMS OF RNA-MEDIATED DNA REPAIR AND MODIFICATION

A Dissertation
Presented to
The Academic Faculty

by

Chance Meers

In Partial Fulfillment
of the Requirements for the Degree
Doctorate of Philosophy in the
School of Biological Sciences

Georgia Institute of Technology
May 2020

MOLECULAR MECHANISMS OF RNA-MEDIATED DNA REPAIR AND MODIFICATION

Approved by:

Dr. Francesca Storici, Advisor
School of Biological Sciences
Georgia Institute of Technology

Dr. Loren Williams
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Kirill Lobachev,
School of Biological Sciences
Georgia Institute of Technology

Dr. Anita Corbett
Department of Biology
Emory University

Dr. Frank Rosenzweig
School of Biological Sciences
Georgia Institute of Technology

Date Approved: January 8th 2020

ACKNOWLEDGEMENTS

I want to thank Dr. Francesca Storici for allowing me to be a part of her lab and her support through graduate school. I would also like to thank my committee members for all their advice and support over the years. An additional thanks to Dr. John Graham for his guidance and support transitioning from undergraduate to graduate school. Finally, I appreciate and will remember all the wonderful friends I have made over the years at Georgia Tech and, in particular those of the Storici Lab, Chernoff Lab, Lobachev Lab, Torres Lab and Gaucher Lab.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

AID	Activation-induced Deaminase
AGS	Aicardi-Goutières Syndrome
Alt-EJ	Alternative End Joining
Alt-NHEJ	Alternative Non-homologous End Joining
AP	Apurinic/Apyrimidic
APOBEC	Apolipoprotein B mRNA Editing Enzyme Catalytic Polypeptide 1
BER	Base Excision Repair
ChIP	Chromatin immunoprecipitation
CPDs	Cyclobutane Pyrimidine Dimers
cDNA	Complementary DNA
C-NHEJ	Classical Non-homologous End Joining
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CTD	C-terminal Domain
C-TDR	cDNA templated DSB Repair
D-loops	three-stranded nucleic acid structure, composed of a DNA:DNA hybrid and the associated non-template single-stranded DNA
DSB	Double-Strand Break
DNA	Deoxyribonucleic Acid
dNTP	Deoxynucleoside Triphosphate
dsDNA	Double-Stranded DNA
EMS	Ethyl Methanesulfonate
GFP	Green Fluorescent Protein
GG-NER	Global Genomic Nucleotide Excision Repair

HA	Human Influenza Hemagglutinin
HDR	Homology Directed Repair
HR	Homologous Recombination
HO	Homothallic Switching Endonuclease
H_2O_2	Hydrogen Peroxide
IG	Immunoglobulin
I-SceI	Intron-encoded Endonuclease
IR	Ionizing Radiation
L1	Long Interspersed Nuclear Elements
LCMS	Liquid Chromatography Mass Spectrometry
LINE-1	Long Interspersed Nuclear Elements
LTR	Long Terminal Repeat
MAP	Modified Sequence Alignment Positions
MAT	Mating Type Locus
MMR	Mismatch Repair
MMS	Methyl Methanesulfonate
ncRNA	Noncoding RNA
NER	Nucleotide Excision Repair
NHEJ	Non-homologous End Joining
NTD	N-terminal Domain
OD	Optical Density
ORF	Open Reading Frame
PBS	Primer Binding Site
PCNA	Proliferating Cell Nuclear Antigen
pGAL1	Promoter of the GAL1

POL	Polymerase
(6 – 4) PPs	Pyrimidine (6 – 4) Pyrimidone Photoproducts
pre-mRNA	Precursor mRNA
pTEF	Promoter of TEF
PTM	Post Translation Modification
RER	Ribonucleotide Excision Repair
R-loops	three-stranded nucleic acid structure, composed of a DNA:RNA hybrid and the associated non-template single-stranded DNA
ROS	Reactive Oxygen Species
RNA	Ribonucleic Acid
RNAi	RNA Interferance
RNAP II	RNA Polymerase II
RNase	Ribonuclease
RNase H	Ribonuclease H
RT	Reverse Transcriptase
R-TDM	RNA Templated DNA Modification
R-TDR	RNA Templated DSB Repair
SAPH-ire	Structural Analysis of PTM Hotspots
SNPs	Single Nucleotide Polymorphisms
SSA	Single Strand Annealing
ssDNA	Single Stranded DNA
ssRNA	Single Stranded RNA
TC-NER	Transcription Coupled Nucleotide Excision Repair
TLS	Translesion Synthesis
Tdt	Terminal Transferase

tRNA	Transfer Ribonucleic Acid
Ty	Transposon of Yeast
UV	Ultra Violet
V(D)J	Variable (V), Diversity (D) and Joining (J)
•OH	Hydroxyl Radicals
($•O2^-$)	Superoxide Radicals
α	Alpha
β	Beta
δ	Delta
η	Eta
ε	Epsilon
γ	Gamma
ι	Iota
κ	Kapa
λ	Lambda
μ	Mu
μJ	Microjoule
μg	Micrograms
ν	Nu
σ	Sigma
θ	Theta
ζ	Zeta

SUMMARY

Maintaining genome stability is a vital aspect for any cellular organism and inability to maintain genome stability can lead to errors during transcription and translation, and if not repaired accurately may give rise to cancerous cells in multicellular organisms. In response, cells developed genetic pathways to combat factors affecting genome instability. One of the most dangerous DNA damaging lesion is a double-stranded break (DSB). We recently found that RNA can mediate the repair of a DSB. This repair can proceed by a transposon-mediated cDNA amplification of transcript RNA and repair by homologous recombination. In addition, RNA can be used directly as a template for repair in the absence of ribonuclease H (RNase H). We have characterized factors of the DNA repair pathways that facilitate RNA-templated DSB repair including *RAD52* and the DNA Polymerase ζ translesion synthesis pathway. Finally, we show that RNA can modify genomic DNA in the absence of an induced DSB in a manner that is independent of *RAD52*, illustrating a rare *RAD52*-independent recombination pathway. This work highlights a complex interplay between DNA and RNA in genome stability.

CHAPTER 1. INTRODUCTION

1.1 DNA Damage and Cellular Responses

Genetic information is coded in the form of deoxyribonucleic acid (DNA) and this information is transferred from parent to offspring to provide instructions for the development of that offspring. Cellular organisms have evolved complex relationships between DNA, ribonucleic acid (RNA) and proteins to carry out all the cellular functions required for life to survive and propagate. As DNA is the information storage molecule of life maintaining the integrity of DNA is critical for accurate storage and transmission. However, DNA is continuously compromised by a variety of different stresses leading to DNA damage. In response, cells have evolved mechanisms adapting DNA, RNA and protein macromolecules to aid in the response to DNA damage.

1.1.1 Exogenous DNA Damage

Life on earth is faced with a plethora of environmental challenges that must be sensed and responded to, including exogenous agents that may alter or damage the genetic code. Ionizing Radiation (IR) consist of alpha, beta, gamma, neutrons and X-rays. IR can damage DNA both directly and indirectly. Indirect IR is attributed to radiolysis of water inside the cell to generate clusters of highly reactive hydroxyl radicals (•OH) termed reactive oxygen species (ROS) [1]. These ROS account for about 65% of the radiation-induced DNA damage [2]. ROS can generate a large variety of damaging events including approximately 100 different oxidative lesions and 2-deoxyribose modification with major lesions consisting of 8-oxo-guanaine, thymine glycol and formamidopyrimidines [3].

Furthermore, IR creates single-stranded breaks resulting in 3′ phosphate or 3′-phosphoglycolate ends instead of 3′-OH ends [4]. If this damage occurs in close proximity, a double strand break (DSB) could occur.

UV radiation discharged from the sun is another source of exogenous DNA damaging agent and the leading cause of skin cancer in humans [5]. UV radiation is classified by the range of wavelength: UV-C (190–290 nm), UV-B (290–320 nm) and UV-A (320–400nm). DNA damage resulting from UV radiation is primary the result of UV-C and can lead to covalent linkages between two adjacent pyrimidines with the two main products being cyclobutane pyrimidine dimers (CPDs) and pyrimidine (6 – 4) pyrimidone photoproducts ((6 – 4) PPs [3, 5].

Exogenous chemical agents also threaten genome stability through a variety of different mechanisms. Exogenous alkylating agents often associated with tobacco smoke and the burning of other biomass creates electrophilic molecules that are highly reactive with the nucleophilic base ring of nitrogen in DNA bases, in particular the N7 of guanine and N3 of adenine. Other commonly known alkylating agents include methyl methanesulfonate (MMS), ethyl methanesulfonate (EMS), and sulfur and nitrogen mustards [6]. Other exogenous sources include aromatic amines and polycyclic aromatic hydrocarbons with two or more aromatic rings, often found in high temperature cooking of foods. The aromatic compounds are processed to produce intermediates that can react with DNA by a variety of mechanisms. One well studied example is benzo(a)pyrene, whose reactive intermediates are known to intercalate into DNA to form DNA adducts [7]. Other exogenous DNA damaging agents include environmental stresses such as extreme heat or cold, hypoxia, and oxidative stress [3].

1.1.2 Endogenous DNA Damaging Agents

In addition to exogenous stresses modifying DNA, endogenous cellular processes can alter the genetic code. The human genome contains over 3 billion base pairs that must be faithfully reproduced every time cells divide. DNA replication is a highly choreographed process accomplished by high-fidelity replicative DNA polymerases (δ and ε). However, cells contain a multitude of specialized DNA polymerases as illustrated by the plethora of human DNA polymerases (α, β, σ, γ, λ, REV1, ζ, η, ι, κ, θ, ν, μ, Tdt and PrimPol), which function to carry out lower fidelity DNA synthesis ty bypass specific DNA lesions [8].

Accurate DNA synthesis activity of DNA polymerases is accomplished by the thermodynamic stability and base-pairing energetics of the incoming dNTP and the template base [9]. Regardless, mistakes do occur with an error rate *in vitro* of about 1 error per 10^4 - 10^5 incorporated nucleotide for Pol α Pol δ and Pol ε [10]. However, mutation rates *in vivo* are much lower, owing to DNA repair processes, described in section 1.1.3. In addition to incorrect base incorporation, DNA polymerases can also incorporate structurally similar rNTPs instead of dNTPs, which contain a 2'OH on the sugar molecule of the nucleotide. Incorporation of ribonucleotides into DNA results in cleavage of ribonucleotide on the 5' end to initiate removal and repair in a process known as ribonucleotide excision repair (RER) [11].

Another source of endogenous DNA damage is the result of topoisomerase enzymes designed to remove super helical tension on DNA during replication and transcription [12]. One of the most commonly studied topoisomerase enzymes is Top1,

which functions to transiently nick DNA and facilitate rotation of the broken strand around the Top1-bound DNA strand and relegation of the broken ends [13]. Although, Top1-DNA complexes can become stabilized and trapped, which can be resolved though reversal of the Top1 complex or excision by DNA endonucleases resulting in DNA damage [14].

One of the most diverse types of endogenous DNA damage is base modification. Base deamination of cytosine, adenine, guanine and 5-methylcytosine results in uracil, hypoxanthine, xanthine and thymine, respectively. These modifications can alter base pairing during DNA replication and lead to mutations. However, certain deamination products such as cytosine deamination to uracil is recognized by uracil-DNA glycosylate [15] and removed by the base excision pathway described in 1.1.3. Cellular deaminases AID (activation-induced deaminase) and APOBEC (Apolipoprotein B mRNA editing enzyme catalytic polypeptide 1) are thought to function as a source of somatic hypermutagenesis in antibody development [16] and host defense mechanisms against retroviruses [17], respectively. Another form of base modification is the generation of abasic or AP (apurinic/apyrimidic) sites. Abasic sites can result from cleavage of the N-glycosyl bond linking the base to the sugar phosphate backbone by spontaneously hydrolysing or by DNA glycosylase cleavage. The resulting abasic sites are removed by AP endonucleases that cleave the 5' end of the abasic site and allow for repair by the base excision repair (BER) pathway [18] or bypass by the translesion synthesis pathway described in section 1.1.3 [19].

As a by-product of the electron transport chain during aerobic respiration, ROS is produced endogenously [20]. As described in section 1.1.1, ROS produces wide variety of DNA damages. The most prominent forms of ROS include superoxide radicals (•O2−),

hydrogen peroxide (H_2O_2), and the hydroxyl radical (•OH). To protect against endogenous ROS, cells restrict respiration to mitochondria and utilize anti-oxidant enzymes to quench surplus ROS [21].

1.1.3 DNA damage response to damaged DNA bases

To protect against the threats of DNA damage, cells developed pathways to combat this damage. These pathways are complex cellular networks to sense and repair altered/damaged DNA sequences. DNA damages may be characterized into two types, those effecting the nucleoside of DNA (base and/or sugar molecule) or those impacting the phosphate backbone linkage. Specialized pathways exist for both types of damages. Alterations in the nucleoside of DNA can lead to mis-incorporation of the wrong nucleotide, polymerase stalling or disruption of the phosphate backbone linkage as part of the repair process. Furthermore, damages in phosphate backbone linkage inhibits cellular division and can ultimately lead to cell death if not corrected.

DNA mismatch repair (MMR) is a system of repairing inaccurate insertions, deletions and mis-incorporation of bases. Mismatch repair is a conserved process from prokaryotes to eukaryotes. Proteins involved in mismatch repair are referred to as "Mut" and cause hypermutability when inactivated [22]. The core components of mismatch include MutS, MutH and MutL in prokaryotes. In eukaryotic cells, these proteins are referred to as MutSα (Msh2/Msh6) and MutSβ (Msh2/Msh3), and MutL homologs include MLH1, MLH2, MLH3, PMS1, and PMS2. Eukaryotic cells contain no homolog of MutH [23]. MutS recognizes mismatched base (MutSα) or small insertions or deletions (MutSβ) in newly synthesized DNA and nicking of the DNA occurs by MutH endonuclease in some

prokaryotic cells or MutL homologs in eukaryotic cells. The nicked DNA is then used as a substrate for exonuclease RecJ or Exo1 and removal of the mismatch site and surrounding nucleotides creating ssDNA that can then be filled in by DNA polymerase [24].

Base excision repair (BER) pathways sense single base damages that are not recognized as distortions to the DNA helix. BER initiation begins with recognition of DNA lesion recognition by a DNA glycosylase. Eleven distinct mammalian DNA glycosylases are known, each of which recognizes different lesion but with some overlapping specificities [25]. Upon recognition of DNA lesion, DNA glycosylases excise damaged bases resulting in abasic sites. The abasic sites are then recognized by the AP endonuclease, cleaving the phosphodiester bond 5′ to the abasic site generating a hydroxyl residue at the 3′-end that can be extended by gap filling synthesis of DNA polymerase. The filling of the gap may proceed by short patch repair of the abasic site or long gap repair requiring more extensive synthesis and removal of 5' flap by the flap endonuclease (Fen1) [26].

Nucleotide excision repair (NER) is used to repair bulky DNA lesions that are not recognized by specific DNA glycosylases in BER. NER can recognize a wide variety of DNA damage that thermodynamically destabilizes the DNA duplex. NER is particularly important for the removal of damaged induced by ultraviolet light including thymine dimers and 6,4-photoproduct (UV). Mutations in the NER pathways are associated with xeroderma pigmentosum, cockayne syndrome and trichothiodystrophy [27]. For DNA damage repaired by NER, the damage is recognized in a transcription coupled manner (TC-NER) or in a global genomic manner (GG-NER). DNA damage is recognized by XPC-RAD23B in (GG-NER) vs in (TC-NER), which is coupled to stalling of RNA polymerase at sites of DNA damage aided by TC-NER specific factors CSA, CSB, and XAB2 [27].

After recognition of lesions in either GG-NER or TC-NER, XPF is recruited to nick 5' to the damage site, followed by 3' incision by XPG removing the damaged site and creating ssDNA gap that is filled in with DNA polymerase [3].

Translesion DNA synthesis (TLS) is orchestrated by specialized TLS polymerases. TLS polymerases include 11 known polymerases REV1, POL η, POL ι, POL κ, POL ζ, POL μ, POL λ, POL β, POL ν, POL θ [8]. These TLS polymerases are unique in their ability to replicate opposite of aberrant DNA damage, although at a much lower fidelity relative to the replicative polymerase (α, δ and ε). These are distributed in four polymerase families (Y, B, X and A) based on structural homology. A model of translesion polymerase synthesis suggest stalling of replication polymerases at the replication fork results in switch of replicative polymerase to translesion polymerases, which mediates the bypass of the damaged DNA.

1.1.4 DNA double strand break repair

DNA double-strand breaks (DSB) are one of the most dangerous lesions encountered as this inhibits a cell ability to replicate its DNA and divide. DSB repair can proceed by two mechanisms: 1) non homologous end joining (NHEJ) and 2) homologous recombination (HR). NHEJ proceeds by initial recognition of DSB by the Ku heterodimer to prevent end resection, which aids in recruiting others NHEJ factors like DNA-PKcs, XRCC4, LIG4 and XRCC4. These factors help initiate end processing which may involve removing damages blocking the DNA ends from repair. This end processing can leave gaps that need to be filled in by DNA polymerase, but this repair process often leaves short insertions or deletions at the break site [28]. Alternatively, HR involves the exchange of

genetic information from homologous sequences. Following a DSB, resection at the site of the DSB generates 3’ overhangs than can be coated by the Rad51 recombinase with the aid of Rad52 and RPA. Rad51 recombinase probes the genome for sequences of homology with the 3’ end. This 3’ end can then be extended by DNA polymerase, transferring information from a homologous sequence to the site of a DSB, fulfilling accurate repair of the DSB [29]. In addition, pathways of single strand annealing and microhomology mediated end joining can proceed with limited homology to sequences around the DSB site, which can result in insertions or deletions [29]. These mechanisms highlight a robust cellular response to DNA DSBs.

1.1.5 RNA and its role in DSB repair

RNA molecules are abundant in cells with estimates as high as 90% of the genome being transcribed [30]. However, RNAs role in the DNA damage response is still largely uncharacterized. RNA is considered to be a driver of genome instability, as base pairing of RNA molecules with DNA termed R-loops, can block processes such as transcription and replication, but also displace the non-complementary DNA strand. This non-complementary strand of ssDNA is then susceptible to DNA damage [31]. However, highly conserved RNase H enzymes target R-loops and RNA-DNA hybrid structures, degrading the RNA strand hybridized to DNA [32]. Regardless, these hybrids can still form at DSB sites either by *de novo* transcription at the break site [33] or by annealing of nascent transcripts. The role of RNA-DNA hybrids at sites of DNA damage is unclear with conflicting reports suggesting both inhibition [34] and stimulation of the DNA damage response [33, 35]. In addition, site specific non-coding small RNAs processed by parts of the RNAi pathway are hypothesized to modulate the DNA damage response [36].

Direct roles for RNA in templating DSB repair is vague, given RNA is largely homologous to the DNA from which it is generated from. Early work in the transposon field showed that LTR (Long Terminal Repeat) containing retrotransposons of yeast (Ty) can insert Ty cDNA at sites of DSBs to repair broken chromosomes but this process is homology-independent and reliant of NHEJ factors [37]. Similarly, human non-LTR retrotransposons, LINE-1, which transpose through a target-primed reverse transcription model, can target LINE-1 integrations at DSBs [38]. In addition, both LTR retrotransposons of yeast (Ty) and Non-LTR retrotransposons of humans (LINE-1) have been shown to retain ability to reverse transcribe cellular RNAs [39, 40]. These reverse transcribed cellular RNAs can mediate recombination, leading to the creation of pseudogenes. Nonetheless, these processes rely on cDNA intermediates to template repair of DSB rather than RNA directly.

An RNA-templated mechanism has been proposed to drive somatic hypermutation at immunoglobulin (Ig) genes [41]. In addition, long RNA molecules can guide large scale genome rearrangements in ciliates [42-44]. Similarly, RNA or RNA-containing oligonucleotides can mediate gene targeting in yeast [45], *E. coli* and human cells [46]. We recently developed an *in vivo* genetic assay to detect RNA-templated DSB repair in budding yeast. We found that in this genetic assay, RNA-templated DSB repair required deletion of RNase H enzymes, as repair in wild-type cells was dependent on the Ty transposon of yeast to mediate DSB repair by a cDNA intermediate [40]. Furthermore, RNA-templated DSB repair strongly requires Rad52, which was shown to catalyze annealing between RNA and DNA *in vitro* [40, 47]. However, the mechanism of RNA-

templated DSB repair and how transcript-RNA interacts with other DNA repair pathways is largely unknown.

1.1.6 Research goals

Understanding the impact of RNA on DNA repair is difficult given that DNA templates the creation of RNA molecules in a cell. We aim to explore the mechanisms by which RNA repairs and/or modifies DNA to better characterize the impact that RNA has on genome stability.

1.1.6.1 To study the impact of RNase H and Ty retrotransposon on RNA-DNA recombination.

We previously showed that RNA can mediate the repair of a DSB in the absence of RNase H in an RNA-templated manner. Loss of functional endogenous retrotransposons of yeast (Ty) by deletion of *spt3* suggested RNA could directly template DSB repair. This process was shown to be more efficient when the donor RNA was generated from its own locus and used for repair vs. a donor RNA transcribed from one locus and used to repair a homologous but ectopic locus. We sought to test that RNA-templated DSB repair is not the result of residual Ty reverse transcriptase in the absence of *spt3* but direct RNA-templated DSB repair.

1.1.6.2 To study reverse transcriptase activity required in the absence of the Ty retrotransposon

We found that RNA-DNA recombination in the absence of RNase H does not require a *bona fide* reverse transcriptase like those seen in retrotransposons. We sought to

investigate how RNA-DNA recombination proceeds in the absence of reverse transcriptase and what cellular polymerases are required for this activity.

1.1.6.3 To study the molecular mechanisms which aid or inhibit RNA-DNA recombination processes

We have previously shown that RNA-DNA recombination proceeds by a homologous recombination mechanism requiring *RAD52*. However, DNA repair factors often have overlapping functions in multiple DNA repair pathways. We choose to investigate the role of other DNA repair pathways including MMR, NER and NHEJ repair factors in RNA-DNA recombination. In addition, we surprisingly found that RNA can modify DNA in the absence of a DSB and sought to characterize these events.

CHAPTER 2. TRANSCRIPT RNA SUPPORTS PRECISE REPAIR OF ITS OWN DNA GENE.

The work in Chapter 2 consists of the work published in *RNA Biology*, *2* (2016), 157-165. I assisted in the writing of the manuscript.

Havva Keskin[1†], Chance Meers[1] and Francesca Storici.[1]

[1]School of Biology, Georgia Institute of Technology, Atlanta, GA 30332

[†]Present address: Omega Biotek, Norcross, Georgia 30332, USA.

2.1 Abstract

The transfer of genetic information from RNA to DNA is considered an extraordinary process in molecular biology. Despite the fact that cells transcribe abundant amount of RNA with a wide range of functions, it has been difficult to uncover whether RNA can serve as a template for DNA repair and recombination. An increasing number of experimental evidences suggest a direct role of RNA in DNA modification. Recently, we demonstrated that endogenous transcript RNA can serve as a template to repair a DNA double-strand break (DSB), the most harmful DNA lesion, not only indirectly via formation of a DNA copy (cDNA) intermediate, but also directly in a homology driven mechanism in budding yeast. These results point out that the transfer of genetic information from RNA to DNA is more general than previously thought. We found that transcript RNA is more efficient in repairing a DSB in its own DNA (in *cis*) than in a homologous but ectopic locus (in *trans*). Here, we summarize current knowledge about the process of RNA-driven DNA repair and recombination, and provide further data in support of our model of DSB repair by transcript RNA in *cis*. We show that a DSB is precisely repaired predominately by transcript RNA and not by residual cDNA in conditions in which formation of cDNA by reverse transcription is inhibited. Additionally, we demonstrate that defects in ribonuclease (RNase) H stimulate precise DSB repair by homologous RNA or cDNA sequence, and not by homologous DNA sequence carried on a plasmid. These results highlight an antagonistic role of RNase H in RNA-DNA recombination. Ultimately, we discuss several questions that should be addressed to better understand mechanisms and implications of RNA-templated DNA repair and recombination.

2.2 Materials and methods

2.2.1 Yeast strains and plasmids

The background strain used to develop all strains used in this study is the haploid FRO-767 strain (*leu2*::HOcs, matα*Δ*::*hisG*, *ho*Δ*, *hml*Δ::*ADE1*, *hmr*Δ::*ADE1*, *ade1*, *leu2*–3,112, *lys5*, *trp1*::*hisG*, *ura3*–52, *ade3*::GAL::HO) [40]. BDG283 and BDG998 vectors (gifts from D. Garfinkel) were transformed into YS-291, 292 (WT) and YS-486, 487 (*spt3 rnh1 rnh201*) strains. BDG283 contains only *pGAL1* and BDG998 contains the *pGAL1-mhis3-AI* cassette, and both plasmids are centromeric with the *URA3* marker [48]. YCp50pK and phis3.210 vectors are also centromeric with the *URA3* marker. YCp50pK was constructed by cloning a SalI/EcoRI fragment with the kanMX4 gene from pFA6a-kanMX4 plasmid52 into the EcoRI/SalI sites of YCp50.53. To construct the phis3.210 vector, a 210-bp fragment of HIS3 was amplified by PCR from genomic DNA using forward primer 50-ACAGTGCTAAGT-AAGCTTATCTTCCCAGAAAAAGAGGC-

30 (HindIII site underlined) and reverse primer 50-ATTGAGTTCCTA-AAGCTT-TACCACCGCTCTGGAAAGTG-30 (HindIII site underlined). The PCR product was digested with HindIII enzyme and was ligated into the YCp50pK vector, which was also digested with HindIII within the kanMX4 gene. The resulting plasmid was sequenced to

confirm the correct 210-bp *HIS3* insert. Both YCp50pK and phis3.210 were transformed into YS-291, 292 (WT), YS-444, 445 (*rad52*), YS-424, 426 (*rnh1 rnh201*), YS-490, 491 (*rad52 rnh1 rnh201*), YS-440, 441 (*spt3*), and YS-486, 487 (*spt3 rnh1 rnh201*) strains. Genetic methods and standard media were described previously [49].

2.2.2 Fluctuation assay of DSB repair

All strains carrying a plasmid with the *URA3* marker gene were maintained on Ura^- medium. Fluctuation assays of DSB repair at the *his3* locus were done as previously described [40]. Briefly, yeast cells were grown in 50-ml lactose containing medium (YPLac), and incubated for 24 hours at 30 °C. Next day, cells were counted, and 10^7 or 10^8 cells were plated on galactose medium (YPGal) or SC-Ura^-Gal medium. 10^4 cells were also plated on YPGal or SC-Ura^-Gal medium to calculate survival. After 2 d incubation, cells were replica plated on His^- or Ura^-His^- medium, and after 3 d His^+ or Ura^+His^+ colonies were counted. Repair frequency and survival were calculated as previously described [40]. Without galactose induction, no or rare His^+ clones are obtained, as discussed in reference [40].

Table 2.1 Statistical analysis (P-values) of the data.

A	
Genotype of *cis* system	*P*-value
WT + BDG283 *vs.* WT + BDG998	< 0.0001
WT + BDG283 *vs. spt3 rnh1 rnh201* + BDG283	< 0.0001
WT + BDG998 *vs. spt3 mh1 mh201* + BDG998	< 0.0001
spt3 rnh1 rnh201 + BDG283 *vs. spt3 mh1 mh201* + BDG998	0.4356
B	
WT + YCp50pK *vs.* WT + phis3.210	< 0.0001
WT + YCp50pK *vs. rnh1 rnh201* + YCp50pK	< 0.0001
WT + YCp50pK *vs. spt3 rnh1 mh201* + YCp50pK	< 0.0001
WT + YCp50pK *vs. rnh1 rnh201 rad52* + YCp50pK	< 0.0001
WT + phis3.210 *vs. spt3* + phis3.210	0.5425
WT + phis3.210 *vs. rnh1 rnh201* + phis3.210	< 0.0001
WT + phis3.210 *vs. spt3 mh1 mh201* + phis3.210	0.1938
WT + phis3.210 *vs. rnh1 rnh201 rad52* + phis3.210	0.0001
rnh1 mh201 + YCp50pK *vs. mh1 mh201* + phis3.210	< 0.0001
rnh1 mh201 + YCp50pK *vs. spt3 rnh1 rnh201* + YCp50pK	< 0.0001
rnh1 mh201 + YCp50pK *vs. mh1 mh201 rad52* + YCp50pK	< 0.0001
rnh1 mh201 + phis3.210 *vs. spt3 rnh1 rnh201* + phis3.210	< 0.0001
rnh1 mh201 + phis3.210 *vs. mh1 mh201 rad52* + phis3.210	0.0001
spt3 + phis3.210 *vs. mh1 mh201* + phis3.210	0.0012
spt3 + phis3.210 *vs. mh1 mh201 rad52* + phis3.210	0.0004
spt3 + phis3.210 *vs. mh1 mh201 spt3* + phis3.210	0.0830
spt3 rnh1 rnh201 + YCp50pK *vs. rnh1 rnh201 rad52* + YCp50pK	< 0.0001
spt3 rnh1 rnh201 + phis3.210 *vs. rnh1 rnh201 rad52* + phis3.210	0.0001
mh1 mh201 rad52 + YCp50pK *vs. mh1 mh201 rad52* + phis3.210	0.4990

Mann-Whitney U-test was applied to determine whether a statistically significant difference exists between pairs of gene correction frequencies obtained in DSB repair assays. A, Comparison of frequencies presented in Table 2.2. Two groups in a pair were considered to be significantly different when adjusted P-values were less than 0.05. B, Comparison of frequencies presented in Table 2.3. Two groups in a pair were considered to be significantly different when adjusted P-values were less than 0.05.

2.2.3 Data presentation and statistics

Statistical analysis was calculated by using GraphPad Prism 5 (GraphPad Software, La Jolla, CA). Median and 95% confidence limits were expressed for each data sample. Statistical significance differences were calculated by using the nonparametric 2-tailed Mann-Whitney U-test, and all P-values of frequency comparisons are shown in (**Table 2.1)**.

2.2.4 Disclosure of potential conflicts of interest

No potential conflicts of interest were disclosed.

2.3 Results and discussion

2.3.1 Transfer of genetic information from RNA to DNA: Theory and supporting evidence

Can RNA transfer genetic information to DNA beyond the special cases of retroviruses, retrotransposons and telomere synthesis [50, 51]? Can RNA recombine with DNA either directly or indirectly if converted into cDNA? Studies on reverse transcription mediated by retrotransposons of yeast (Tys), or of insects (R2), have shown that not only RNA originating from retroelements could be reverse transcribed but potentially any RNA [52], such as the RNA deriving from the yeast *HIS3* marker gene, and that RNA could mediate recombination with DNA and modify genomic DNA once converted into cDNA via reverse transcription [48, 53, 54]. It was found that not only Ty cDNA, but also HIS3 cDNA could recombine with homologous or homeologous (partially homologous) DNA [54], integrate into genomic DNA if fused to transposon sequences, or be captured at sites of chromosomal DSBs via non-homologous end joining (NHEJ) [55, 56]. Additional studies in yeast revealed involvement of cDNA in homologous recombination (HR) [57-

59], and it was suggested that different types of reverse transcription products including ssDNA and RNA-DNA hybrids could be engaged in recombination [60]. Further work in mammalian cells showed that Long INterspersed Elements (LINEs) can be captured at sites of DNA damage, and that retrotransposition of LINEs can carry fragments at their ends that are derived from reverse transcription of endogenous mRNA [38, 39, 61].

There has been a series of hypotheses and speculations that RNA can work as a template in DNA recombination and repair [62]. Recombination mediated by reverse transcripts of cellular RNAs with homologous DNA has been suggested to explain the paucity of introns in yeast genomic DNA, while end-joining-driven insertions of cDNA products could explain the abundance of pseudogenes in multicellular eukaryotes [63, 64]. Indeed mRNA-mediated intron losses were shown to occur in yeast mitochondrial DNA [65] (and references therein). Murakami et al. suggested a mechanism of RNA-directed DNA repair in mitochondria facilitated by the reverse transcriptase activity of DNA polymerase gamma [66], whereas possible mechanisms of DSB repair in nuclear DNA by RNA have been proposed by Trott and Porter [67]. The discovery of a widespread type of viral genome representing a chimera between an RNA and a DNA virus has inferred the occurrence of RNA-DNA recombination between two quite different virus groups [68, 69].

From work in plants, Xu et al. proposed a direct or indirect RNA-templated DSB repair mechanism via gene conversion to explain the observed high frequency of gene homozygosity in rice [70]. Furthermore, a recent study reported that DSBs in neurons of young adult mice can be part of normal brain functions such as learning, as long as the DSBs are controlled and repaired in short time [71]. Could RNA serve as template for DNA repair of these physiological DSBs in neurons? It has been proposed that flow of

information from RNA to DNA could lead to DNA recoding events in the nervous system and could be the basis for permanent storage of long term memories [72, 73].

Considering the abundance of RNA in cells, the flow of genetic information from RNA to DNA could strongly affect genome stability, either by increasing or decreasing it, depending on the circumstances. Different experimental insights suggest that mechanisms of RNA-driven DNA modification might be more common than is currently recognized. Evidence of RNA-derived insertions came from analysis of sequences at DSB sites in fruit fly and mammalian cells. An exon–exon junction sequence was found from the analysis of DNA sequences repaired via NHEJ after DSB induction by zinc-finger nucleases in Drosophila cells [74], suggesting a direct or indirect RNA-templated insertion mechanism. Work in human cells revealed presence of murine sequences derived from murine RNA that was co-transfected into the human cells together with the DNA of the I-SceI DSB-inducing vector [75]. More recently, exonic RNA insertions were detected in knock-in mouse experiments at sites of DNA DSBs generated using the CRISPR/Cas9 system [76]. Overall, these studies showed that insertions of RNA derived sequences can result in an error-prone form of DNA repair, which may play a role in genetic disorders and evolution.

2.3.2 Is there experimental proof for RNA-DNA recombination and RNA-mediated DNA repair that is homology driven?

Can RNA directly mediate genetic DNA modifications in a homology-driven manner? Can RNA repair a DSB in homologous DNA sequences? Experiments in budding yeast showed that not only short ribonucleotide tracts carried within synthetic DNA oligonucleotides (oligos) but also RNA-only oligos can precisely repair a DSB in

homologous DNA, serving as direct templates for DNA synthesis at the chromosomal level, and transferring genetic information also in conditions in which Ty reverse transcription is repressed [45, 77, 78]. The capacity of short RNA patches to directly modify DNA was also found in the bacterium *Escherichia coli* [79, 80], and RNA oligos could precisely repair a DSB in the green fluorescent protein gene in human embryonic kidney (HEK-293) cells [80].

As a model to explain the occurrence of transgenerational inheritance of genomic DNA rearrangements in ciliated protozoa, Angeleska et al. proposed a mechanism in which RNA molecules, single- or double-stranded (ss or ds), act as template catalyst to guide specific recombination events [81]. The model for RNA-templated DNA rearrangements was then tested using long synthetic RNA sequences injected into the ciliate *Oxytricha trifallax* and the RNA templates were found to mediate correct and precise DNA rearrangements [43, 44]. In addition, mutations carried on the artificial RNA templates were transferred to the homologous endogenous DNA sequences suggesting a process of RNA-guided DNA repair in *O. trifallax* [44].

Models of RNA-DNA HR are supported by biochemical studies, showing the ability of the *E. coli* recombinase RecA to promote pairing between duplex DNA and ssRNA *in vitro* [82-85]. Moreover, recent work suggests that the eukaryotic RecA homolog, Rad51, can also promote formation of RNA-DNA hybrids in yeast [86].

Beyond the demonstration that synthetic RNA molecules introduced into cells can mediate HR with DNA, our recent work showed that endogenous transcript RNA can be a template for DSB repair and HR in yeast [40]. We provided experimental evidence that the

transfer of genetic information from RNA to DNA occurs with an endogenous generic transcript, and is thus a broader phenomenon than previously anticipated.

2.3.3 Transcript RNA mediates DSB repair in a homology-driven manner

We developed a system to explore the prospects of an endogenous RNA transcript ability to serve as a template for the repair of DSBs, casting a new light on the roles of RNA in the DNA damage response [40]. Our strategy is based on the induction of a DSB located inside a nonfunctional *his3* marker gene, and successive DSB repair via an endogenous spliced transcript RNA resulting in histidine prototrophic (His^+) cells. We engineered *cis* and *trans* systems granting the possibility to evaluate the effects of localization and continuous productions of the transcript RNA. The *cis* system transcribes an antisense *his3* sequence with an artificial intron inserted in the antisense orientation that upon galactose induction results in a spliced antisense *his3* transcript that can facilitate repair of a DSB located inside the artificial intron resulting in a functional *HIS3* locus. The artificial intron (105 bp) contains the site for the HO endonuclease (124 bp); in total, a 229-bp insert disrupts the *HIS3* gene. Likewise, the trans system is based on dual *his3* loci, in which, one locus is the endogenous *HIS3* gene on chromosome XV but disrupted by the cutting site of the HO endonuclease, and the other locus is located on chromosome III and serves to produce an antisense *his3* transcript with an artificial intron inserted in the antisense orientation that upon galactose induction produces a *his3* antisense transcript that can aid in the repair of the DSB generated at the HO site of the endogenous *HIS3* (**Figure 2.1**). Considering the abundance of retrotransposons in the yeast genome [87], we sought to eliminate the reverse transcription activities associated with retroelements to explore the ability of RNA to serve directly as a template for repair rather than through the cDNA

intermediates of retrotransposition. To this end, we created an *spt3*-null mutant, which prevents normal Ty transcription and reduces Ty transposition [88]. As a result, in *spt3*-null mutant yeast, no His^+ colonies are observed suggesting that cDNA-mediated repair is the major pathway of repair in transposition proficient cells [40]. This indicates that any actively transcribed gene can be repaired using a reverse transcribed cDNA template. Because an RNA-DNA heteroduplex is a probable prerequisite for RNA to recombine directly with DNA, we sought to facilitate stable formation of RNA-DNA hybrids by deletion of RNase H1 (*RNH1*) and the catalytic subunit of RNase H2 (*RNH201*) genes, which both code for nonsequence-specific endonucleases that cleave RNA backbone of RNA-DNA hybrids [89]. Deletion of both *RNH1* and *RNH201* results in a 5-fold increase of His^+ colonies in *trans* and a 35-fold increase in *cis*. Surprisingly, the *spt3 rnh1 rnh201* genotype results in more than 69,000 His^+ colonies than in *spt3* single mutant, and even more intriguingly, the *cis* system of the *spt3 rnh1 rnh201* cells yields 10-fold more His^+ colonies than the *trans* system, which continuously produces transcript for repair [40]. Furthermore, deletion of the *RAD52* gene, which codes for an important homologous recombination protein facilitating the annealing of complementary ssDNA, results in a strong reduction of His^+ colonies in *spt3 rnh1 rnh201* cells [40]. A complementary *in vitro* study suggests that yeast and human *Rad52* can promote the annealing of RNA to DNA, and in the presence of RPA, even more efficiently than DNA to DNA [40]. Thus, we propose a model that upon the occurrence of a DSB in a transcribed DNA, *Rad52* promotes the annealing of RNA to DNA, and, in the absence of RNases H, RNA serves as a template bridging the broken DNA ends to promote precise re-ligation, or allowing extension of the broken end via reverse transcription [40]. Given the prerequisite that our assay requires a

spliced mRNA to display a phenotype, we could be missing repair by unspliced mRNA, thus RNA-templated DNA modifications may have a substantial impact on genomic stability.

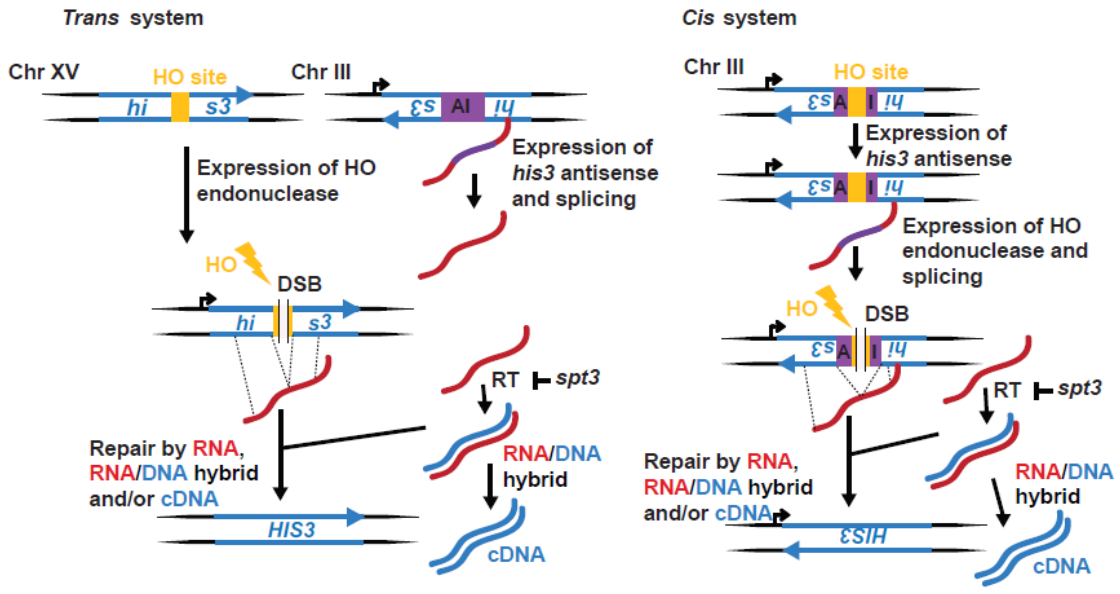


Figure 2.1 Scheme of the trans and cis systems.

HO, homothallic switching endonuclease (yellow); AI, artificial intron (purple); right turn arrow, *pGAL1*; yellow lightning bolt, cleavage activity by HO; RT, reverse transcriptase.

2.3.4 DNA self-repair by transcript RNA

Our results of DSB repair in the *cis* and *trans* systems showed that the frequency of His^+ colonies in *cis spt3 rnh1 rnh201* cells was >69,000-fold higher than in *cis spt3* cells, and >10 –fold higher than in *trans spt3 rnh1 rnh201* cells [40]. Is this high frequency of His^+ colonies in *cis spt3 rnh1 rnh201* cells due to the RNA functioning as homologous template to mediate a precise re-ligation of the broken DSB ends? Alternatively, is this repair templated by cDNA due to residual Ty activity? We showed that the DSB repair at

the *his3* locus in *cis spt3 rnh1 rnh201* cells was predominately mediated by transcript RNA rather than cDNA [40]. Here, we corroborate our finding that transcript RNA can directly serve as a template for repair of a DSB occurring in the same DNA that generated the transcript *in spt3 rnh1 rnh201* cells of the *cis* system.

We examined the effect of an extra copy of the *his3* allele, disrupted by the artificial intron in the antisense orientation (*mhis3-AI*) carried on a yeast centromeric plasmid (BDG998) (**Figure 2.2A**), on the frequency of His^+ colonies following DSB induction in wild-type and *spt3 rnh1 rnh201* backgrounds of the *cis* system. We transformed wild-type and *spt3 rnh1 rnh201* cells with low copy number plasmid BDG998 or with the control empty plasmid (BDG283), which carry the *URA3* marker gene (**Figure 2.2A**) and selected for colonies able to grow on medium lacking uracil (Ura^+ colonies). We then performed the fluctuation assay as described in Materials and Methods and in [40]. In wild-type cells, the His^+ frequency was strongly increased in the presence of the BDG998 plasmid (**Table 2.2 and Figure 2.3**) compared to BDG283-containing cells. This was expected because not only the *his3* antisense transcribed from the chromosomal *his3* copy, but also the one transcribed from the *his3* copy carried on the BDG998 plasmid can be converted into cDNA by Ty reverse transcriptase and provide additional copies for DSB repair. Moreover, differently from the chromosomal copy, the plasmid copy of *his3* can continue to be transcribed in galactose medium because it does not contain the site for the HO endonuclease within the artificial intron, thus, it can generate lots of cDNA molecules. In contrast, there is no significant difference in the frequency of His^+ colonies between *spt3 rnh1 rnh201* cells containing BDG283 and BDG998 (**Table 2.1 and Figure 2.3**). If cDNA would be the major template for *his3* repair in *spt3 rnh1 rnh201* cells we would expect

higher frequency of His^+ colonies also when these cells contain BDG998 than in cells containing BDG283. These data suggest that even if there is residual cDNA in *cis spt3 rnh1 rnh201* cells, cDNA does not play a major role in DSB repair of the *his3* locus. Rather, it is the transcript RNA from the chromosomal locus that mediates, in *cis*, most of DSB repair to restore the function of its broken *his3* gene on the chromosome.

2.3.5 Defects in RNase H activity stimulate homology-driven DSB repair by cDNA and RNA, but not by plasmid DNA

Our findings show that absence of RNase H1 and/or H2 activity in wild-type or null-*spt3* cells results in increased frequency of His^+ colonies after DSB induction not only in the *cis* but also in the *trans* system compared to wild-type RNase H cells [40]. These results indicate that absence of RNase H function activates DSB repair by transcript RNA, and also stimulates DSB repair by cDNA. Following reverse transcription of RNA into cDNA, cDNA can be present as RNA-DNA hybrid, ssDNA, and/or dsDNA. Previously, we showed that DSB repair by ssDNA oligos was not increased in *rnh1 rnh201* cells compared to RNase H wild-type cells [40]. Moreover, our recent work indicates that defective RNase H2 alleles have higher level of cDNA in the form of RNA-cDNA hybrids [90]. Here, we examined whether the RNase H defect is specific to stimulate DSB repair of the broken *his3* locus via HR only by RNA and/or cDNA, or it can also stimulate DSB repair by gene conversion using as template for HR a truncated *his3* copy carried on a dsDNA plasmid. We transformed wild-type, *rad52*, *rnh1 rnh201*, *rnh1 rnh201 rad52*, *spt3* and *spt3 rnh1 rnh201* strains of the *cis* system with a plasmid carrying an internal 210-bp segment of the *HIS3* gene sequence (phis3.210) or with the control empty plasmid (YCp50pK) (**Figure 2.2B**). To determine the frequency of His^+ colonies following DSB

induction at the *his3* chromosomal locus for all these strains, we conducted the fluctuation assay of DSB repair. Depending on the genotype of the strains, cells containing the control vector YCp50pK can repair the DSB in the chromosomal *his3* allele by using as template for HR the RNA, RNA-DNA hybrid and/or cDNA derived from the chromosomal *his3* locus, while cells containing phis3.210, in addition to the RNA, RNA-DNA hybrid and/or cDNA derived from the chromosomal *his3* locus, can also repair the DSB in *his3* by using as template the DNA of the truncated *his3* allele carried on phis3.210 cDNA, and/or potentially the RNA, RNA-DNA hybrid and/or cDNA derived from the transcription of this *his3* plasmid allele (**Table 2.3 and Figure 2.4**). In wild-type cells, there is a factor of 50 increase in the His^+ frequency in the presence of phis3.210 compared to YCp50pK (**Table 2.3**). As expected, upon deletion of the *RAD52* gene, which is required for any mechanism of DNA-DNA HR in yeast [91], no His^+ colonies are detected with either plasmid. In *rnh1 rnh201* cells carrying YCp50pK, the His^+ frequency is more than a factor of 20 higher than in wild-type cells, due to elevated repair by cDNA and RNA, in agreement with our previous findings [40, 87]. The *rnh1 rnh201* mutations in cells carrying phis3.210 result in less than 2-fold increase of the His^+ frequency compared to wildtype cells carrying the same plasmid (**Table 2.3**). Such increase can be explained by the fact that in this background the DSB can be repaired not only by the truncated *his3* on the plasmid, but also by RNA and cDNA derived from the chromosomal *his3* copy, as well as by cDNA derived from the his3 allele on phis3.210. There could also be some repair in *trans* by the RNA derived from the *his3* allele on phis3.210, although we expect this to be minimal compared to repair by cDNA. However, clearly, defects in RNase H1 and H2 do not stimulate DSB repair by the DNA of the *his3* copy on the plasmid. In fact, in *spt3*

mutant cells, in which there is no or very little cDNA, there is no difference in the frequency of His$^+$ colonies between *spt3* and *spt3 rnh1 rnh201* cells carrying phis3.210 (**Tables 2.2 and 2.3**). While there could be some repair in *trans* by the RNA derived from the *his3* allele on phis3.210 in *spt3 rnh1 rnh201* cells, we expect this to be minimal as shown in Keskin *et al*. 2014 [40]. Differently, there is a remarkable difference (more than a factor of 60,000) in the frequency of His$^+$ colonies between *spt3* and *spt3 rnh1 rnh201* cells carrying YCp50pK due to repair by RNA. Deletion of *RAD52* in *rnh1 rnh201* cells prevents repair by the *his3* copy on the plasmid and by cDNA, while, as previously shown [40], it reduces, but not abolishes RNA repair either in the presence of phis3.210 or YCp50pK (**Table 2.3**). Overall, these results demonstrate that absence of RNase H activity does not stimulate DSB repair via DNA-DNA HR, while it strongly activates RNA-DNA HR, and HR between DNA and cDNA, in which the cDNA is most likely an RNA-DNA hybrid.

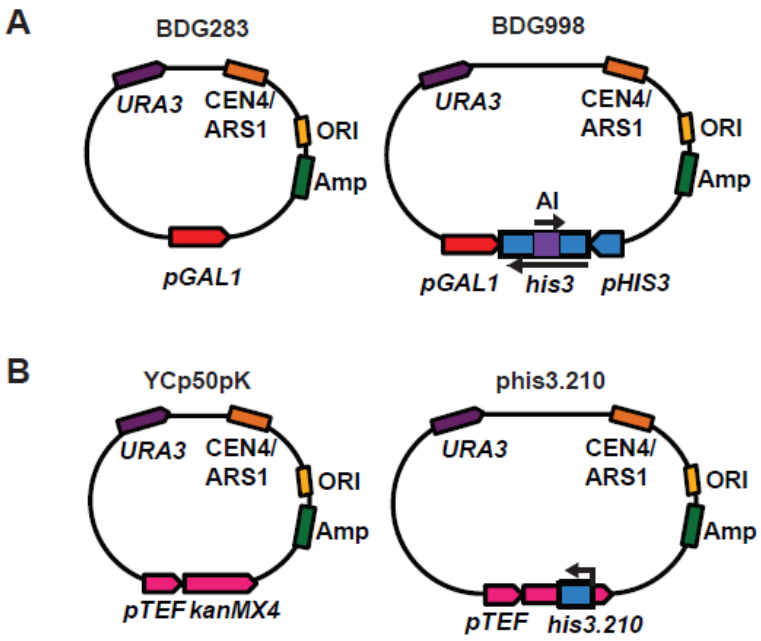


Figure 2.2 Scheme of the plasmids introduced in the *cis* system.

A) BDG283 and BDG998. GAL1 promoter, pGAL1 (red); his3 promoter and open-reading frame, pHIS3 and his3 (blue); AI, artificial intron (purple). The arrows indicate the orientation of the AI and that of the his3 gene. Other parts of the plasmids are also shown. B) YCp50pK and phis3.210. The kanMX4 gene with the pTEF promoter are in pink; 210-bp fragment of HIS3 sequence, his3.210 (blue) is inserted in the kanMX4 gene. The orientation of the his3 fragment is indicated by an arrow. Other parts of the plasmids are also shown.

Table 2.2 Transcript RNA-templated repair is the major mechanism for precise DSB repair in *spt3 rnh1 rnh201* cells in *cis* system.

Genotype of *cis* system	His^+ freq.		Survival
WT + BDG283	33	(20–45)	9%
WT + BDG998	4,130	(2,680–6,190)	10%
spt3 rnh1 rnh201 + BDG283	870	(706–960)	22%
spt3 rnh1 rnh201 + BDG998	890	(850–980)	27%

Frequencies of His^+ colonies per 10^7 viable cells for yeast strains of the *cis* system of the indicated genotypes and containing either the control empty vector BDG283 or vector BDG998, following 48 h of galactose treatment are shown as median and 95% CI (in parentheses). Percentage of cell survival after incubation on galactose is also shown. There were 9–12 repeats for each strain. The significance of comparisons between different strains of the system was calculated using the Mann-Whitney U-test and it is shown in Table 2.1A. Figure 2.3 serves as graphical guide for all results presented in this table.

2.3.6 What's next?

Our recent findings raise a multitude of unanswered questions. We have shown that a transcript RNA can facilitate the repair of a DSB via a direct or indirect cDNA intermediate pathway. What are the players involved in this newly discovered mechanism of DNA repair? What factors mediate the increasing amount of repair in *cis* versus *trans* in *spt3 rnh1 rnh201* cells? Based on the localization of the transcript, nearby its DNA gene, the *cis* system is more prone to the generation of an RNA-DNA hybrid at the *his3* locus. If so, can reverse transcriptase enter the nucleus and facilitate reverse transcription at the site of a DSB? Can other polymerases use RNA as a template in DSB repair in vivo? What is

the real efficiency of transcript-templated DNA repair? Our assay is limited by the detection of a phenotype, His^+ cells, which originate only if the RNA template repairs the DSB after splicing of the artificial intron. If transcript RNA mediates DSB repair before splicing, there is no phenotype detected in our assay. Therefore, it is quite possible that we are underestimating the frequency of DSB repair by template transcript RNA. Does DSB repair by template transcript RNA occur in mammalian cells and in other cell types? We showed that transcript RNA-templated DNA repair occurs in dividing yeast cells. Can RNA template DSB repair in non-dividing cells? For example, highly transcribed genes in non-dividing cells, in which no sister chromatid is available, could be vulnerable; thus, these genes could be liable to RNA-templated DNA repair.

Our results of RNA repairing a DSB indirectly, via cDNA, shed light on the possibility of any RNA molecule being a target for reverse transcription by endogenous retrotransposon activity. If so, what factors mediate this reverse transcription? How abundant is the cDNA generation of endogenous RNA molecules? The Saccharomyces cerevisiae genome contains 5 classes of retroelements known as Tys, with Ty1 being the most abundant and well-studied. Is one class more prone to the generation of cDNA by endogenous RNA molecules? What can these factors tell us about other endogenous retroelements and retroviral infections? Retrotransposons are ubiquitous and plentiful in plant genomes, in some cases accounting for over 50% for the nuclear genome [92]. Mammalian genomes are no strangers to retroelements with »3 million transposable elements in the human genome and 90% of those being retrotransposons [93]. Given the copious amounts of retroelements found throughout various genomes and the relative

abundant amounts of RNA in contrast to DNA, could RNA-templated DNA repair be playing a significant role in genome stability and modification?

Our work has provided fundamental preliminary data and resulted in the development of unique tools to study DNA repair via HR directly by RNA in the yeast model system. While inactivation of RNase H function allowed us to discover the capacity of cells to use transcript RNA in DSB repair, it is possible that RNA-DNA HR occurs also in RNase H wild-type cells. Mechanisms and functions of RNA-DNA HR are mostly unknown. Further studies are needed to illuminate the implications RNA-DNA HR may have on genome integrity.

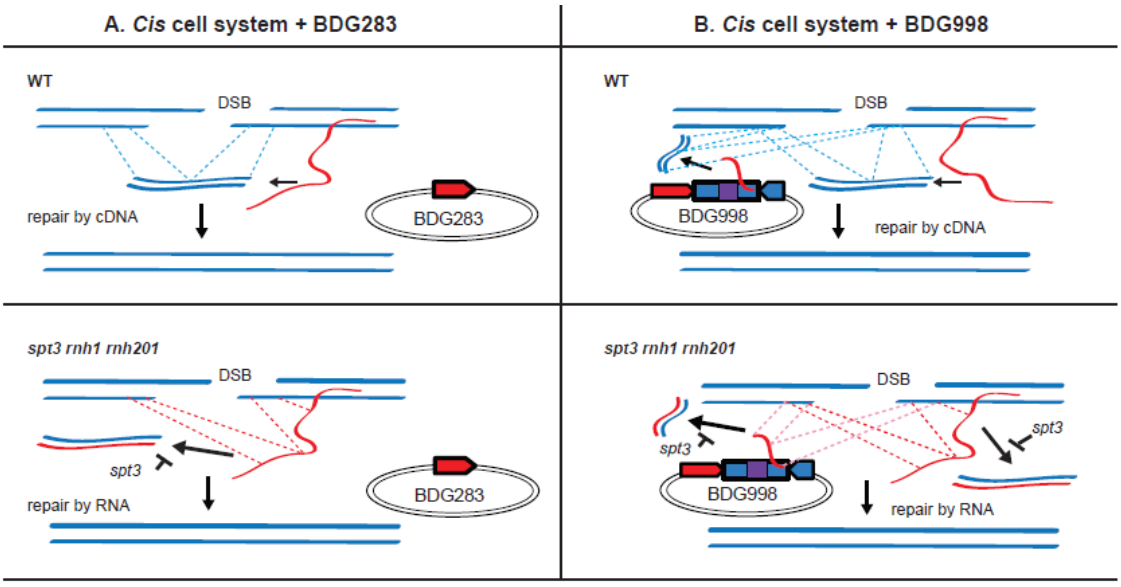


Figure 2.3 Templates for DSB repair in *his3* locus to generate a functional *HIS3* gene in a *trans-cis* system.

This figure reflects the results of Table 2.2. Only repair mechanisms resulting in functional restoration of HIS3 are shown. Repair may also proceed by canonical NHEJ or HR with sister chromatid but does not result in functional HIS3. Regions of homology to the DSB site in his3 are shown as dashed lines. The spt3-null mutation results in inhibition of reverse transcription by Ty retroelements. Relevant genotypes are shown in the top left corner of each panel. Donor molecules that can serve as template for DSB repair are shown as solid blue lines for cDNA and dsDNA, red and blue lines for RNA-DNA hybrid, and red lines

for transcript-RNA. A) Repair of a DSB in cis system in the presence of BDG283. B) Repair of a DSB in cis system in the presence of BDG998. DSB repair in trans templated by the spliced RNA from the transcription of his3 on BDG998 is also possible in cells containing rnh1 rnh201 mutations, although this is inefficient.

Table 2.3 Effect of RNase H1 and H2-null mutations on DSB repair frequency by homologous cDNA, RNA-DNA hybrid, RNA and/or plasmid dsDNA.

Genotype of *cis* system	His^+ freq.		Survival
WT + YCp50pK	2,500	(2,300–2,830)	1.9%
WT + phis3.210	133,000	(104,000–151,000)	2%
rad52 + YCp50pK	<0.1	(0–0)	0.2%
rad52 + phis3.210	<0.1	(0–0)	0.16%
rnh1 rnh201 + YCp50pK	52,300	(47,700–63,300)	1.4%
rnh1 rnh201 + phis3.210	248,000	(226,000–309,000)	1.2%
rnh1 rnh201 rad52 + YCp50pK	930	(790–1,300)	0.07%
rnh1 rnh201 rad52 + phis3.210	1,000	(590–1,200)	0.08%
spt3 + YCp50pK	<0.1	(0–0)	7%
spt3 + phis3.210	134,000	(96,000–180,000)	4%
spt3 rnh1 rnh201 + YCp50pK	6,300	(5,900–7,200)	10%
spt3 rnh1 rnh201 + phis3.210	99,000	(92,000–119,000)	7%

Frequencies of His^+ colonies per 10^7 viable cells for yeast strains of the *cis* system of the indicated genotypes and containing the indicated plasmid, following 48 h of galactose treatment are shown as median and 95% CI (in parentheses). Percentage of cell survival after incubation on galactose is also shown. There were 6–12 repeats for each strain. The significance of comparisons between different strains of the system was calculated using the Mann-Whitney U-test and it is shown in Table 2.1B. Figure 2.4 serves as graphical guide for all results presented in this table.

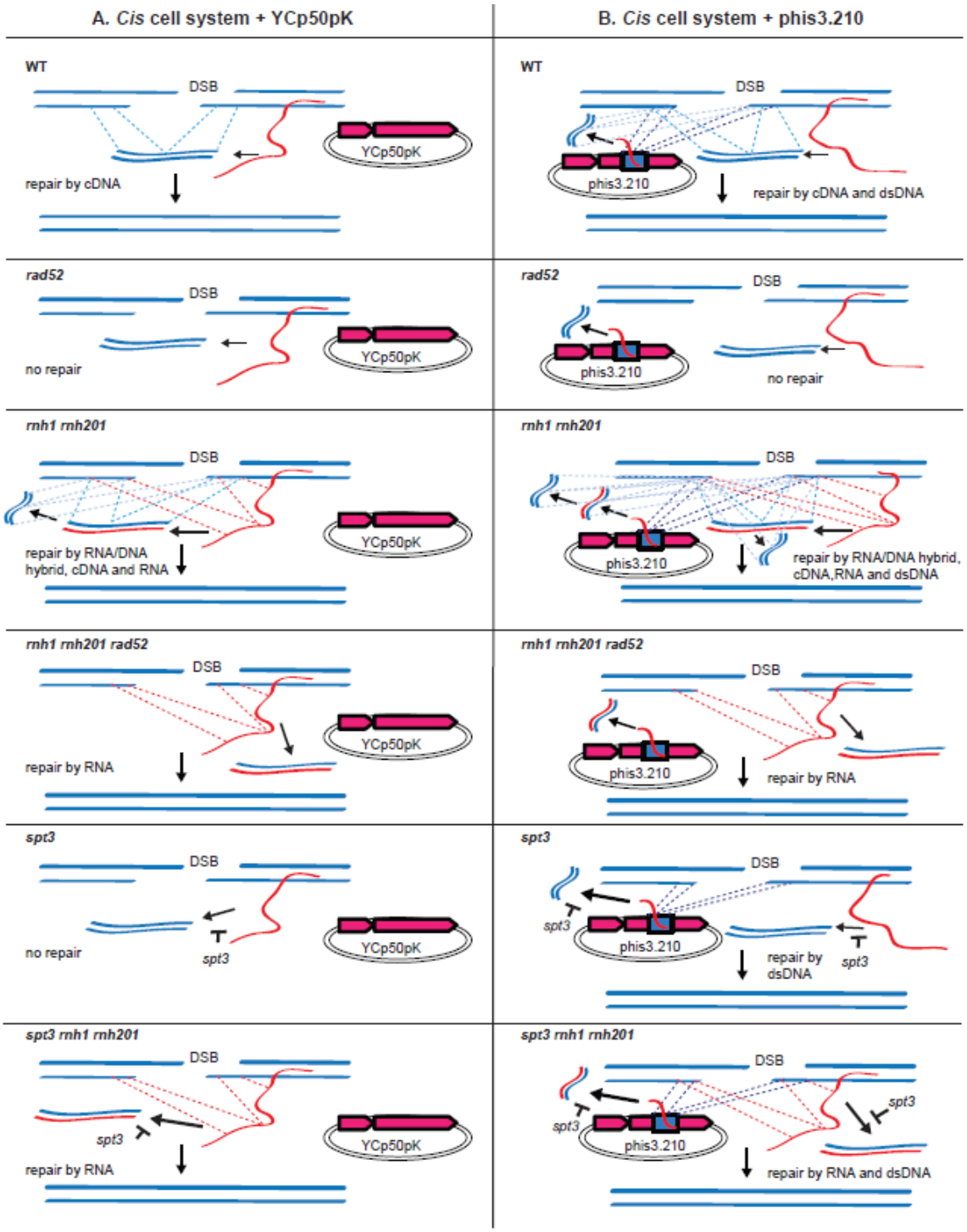


Figure 2.4 Templates for DSB repair in *his3* locus to generate a functional *HIS3* gene in *cis* system.

This figure reflects the results of Table 2.3. Only repair mechanisms resulting in functional restoration of *HIS3* are shown. DSB repair in *his3* may also proceed by canonical NHEJ or HR with sister chromatid but does not result in functional *HIS3*. Regions of homology to

the DSB site in *his3* are shown as dashed lines. The *spt3*-null mutation results in inhibition of reverse transcription by Ty retroelements. Relevant genotypes are shown in the top left corner of each panel. Donor molecules that can serve as template for DSB repair are shown as solid blue lines, for cDNA and dsDNA, red and blue lines for RNA-DNA hybrid, and red lines for transcript-RNA. A) Repair of a DSB in *cis* system in the presence of YCp50pK. B) Repair of a DSB in *cis* system in the presence of phis3.210, which contains 210 bp of *HIS3* (blue rectangle). DSB repair in *trans* templated by the RNA from the transcription of *his3* on phis3.210 is also possible in cells containing *rnh1 rnh201* mutations. Due to inefficient DSB repair by RNA in *trans*, we did not show the dashed lines for this template in the panels.

2.4 Acknowledgments

We thank D. Garfinkel for plasmids BDG283 and BDG998. This study was supported by US National Science Foundation award MCB-1021763, Georgia Research Alliance award number R9028, and National Institute of Health award GM115927 (to F.S.). H.K. was partly supported by a fellowship from the Ministry of Science of Turkey.

CHAPTER 3. DNA REPAIR BY RNA: TEMPLATED, OR NOT TEMPLATED, THAT IS THE QUESTION.

The work in Chapter 3 consists of a review article published in *DNA repair* 44 (2016), 17-21. I devised and wrote the manuscript with input and figures produced by H. Keskin.

Meers, C. [1], Keskin, H. [1†] & Storici, F. [1]

[1]School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA 30332, USA.

[†]Present address: Omega Biotek, Norcross, Georgia 30332, USA.

3.1 Abstract

Cells are continuously exposed to both endogenous and exogenous sources of genomic stress. To maintain chromosome stability, a variety of mechanisms have evolved to cope with the multitude of genetic abnormalities that can arise over the life of a cell. Still, failures to repair these lesions are the driving force of cancers and other degenerative disorders. DNA double-strand breaks (DSBs) are the most toxic genetic lesions, inhibiting cell ability to replicate, and are sites of mutations and chromosomal rearrangements. DSB repair is known to proceed via two major mechanisms: homologous recombination (HR) and non-homologous end joining (NHEJ). HR reliance on the exchange of genetic information between two identical or nearly identical DNA molecules offers increased accuracy. While the preferred substrate for HR in mitotic cells is the sister chromatid, this is limited to the S and G2 phases of the cell cycle. However, abundant amounts of homologous genetic substrate may exist throughout the cell cycle in the form of RNA. Considered an uncommon occurrence, the direct transfer of information from RNA to DNA is thought to be limited to special circumstances. Studies have shown that RNA molecules reverse transcribed into cDNA can be incorporated into DNA at DSB sites via a non-templated mechanism by NHEJ or a templated mechanism by HR. In addition, synthetic RNA molecules can directly template the repair of DSBs in yeast and human cells via an HR mechanism. New work suggests that even endogenous transcript RNA can serve as a homologous template to repair a DSB in chromosomal DNA. In this perspective, we will review and discuss the recent advancements in DSB repair by RNA via non-templated and templated mechanisms. We will provide current findings, models and future challenges investigating RNA and its role in DSB repair.

3.2 Surviving the tragedy of DNA breaks.

The accurate transfer of genetic information is an essential requirement that has allowed life to precisely transmit information from one generation to the next. Over the billions of years DNA based life has existed on Earth, the transfer of genetic materials has continuously been bombarded with exogenous and endogenous threats including ionizing radiation, replication stress, and reactive oxygen and nitrogen species. These threats result in a diverse set of genetic abnormalities including but not limited to modified bases, mispaired bases, cross-linkage, nicked DNA and double-stranded breaks (DSBs). In particular, DSBs in DNA are one of the most severe events a cell can experience. DSBs are hazards to genome integrity, causing mutations, chromosomal rearrangements or cell death if not precisely repaired [94]. To contend with these threats, cells evolved two major repair mechanisms for DSBs: non-homologous end joining (NHEJ) and homologous recombination (HR). NHEJ is an error prone process of reattachment of broken DNA ends, favored in the G0/G1 phase of the cell cycle [95]. In contrast, HR involves the exchange of genetic information between homologous or homeologous DNA sequences and is thus generally active in the S and G2 phases of the cell cycle, allowing for accurate repair by a sister chromatid in mitotic cells [96]. Homologous DNA substrates that facilitate repair are in relative short supply when compared to the abundance of homologous substrate that may be present in the form of transcript RNA. However, the transfer of genetic information from RNA to DNA is considered a rare phenomenon. Could RNA play a significant role in DSB repair?

3.3 Central Dogma Reversed

In the central dogma of molecular biology, the general transfers of genetic information: DNA → DNA, DNA → RNA, and RNA → protein are believed to occur in most cells; while the special transfers: RNA → RNA, RNA → DNA, and DNA → protein are restricted to specific sequences, or have only been shown *in vitro*. The unknown transfers of protein → DNA and protein → RNA were postulated to never take place [97, 98]. The transfer of RNA→ DNA, reverse transcription, is known to happen in retroviruses, retrotransposons, and during telomere synthesis [50, 99]. Remarkably, there is mounting evidence showing that the transfer of RNA → DNA is not merely restricted to these special cases, but is a more general phenomenon in nature. The first evidence for an RNA→ DNA transfer of genetic information not limited to nucleic acid of retroelements came from experiments in the early 1990's on the reverse transcriptase (RT) activity of the yeast retotransposons (Tys). These studies showed that: i) not only is RNA originating from retroelements reverse transcribed in yeast, but potentially any cellular RNA, such as the RNA deriving from the yeast *HIS3* marker gene, may be substrates of reverse transcription, and that ii) RNA reverse transcribed into a DNA copy (cDNA) can modify genomic DNA [40, 48, 90, 100, 101]. The cDNA products are potential substrates for recombination between homologous or homeologous sequences or can be captured at sites of a DSB [54-56]. The work of others on the function of the mammalian retrotransposon Long Intersperse Element-1 (LINE-1) found the capture of endogenous mRNA sequences, in addition to LINE-1 mRNA, at sites of DNA damage [100, 102]. Overall these findings indicate that the transfer of RNA → DNA is not solely limited to special cases and sequences. As discussed below, more recent experimental data have continued to provide support for an RNA→ DNA transfer of genetic information from 'generic' RNA sequences, not deriving

from retroelements or telomeres, even without a cDNA intermediate. Therefore, we propose a broader model in which the transfer of RNA → DNA is also a general type of genetic transfer in the process of RNA-mediated DNA repair (**Figure 3.1)**.

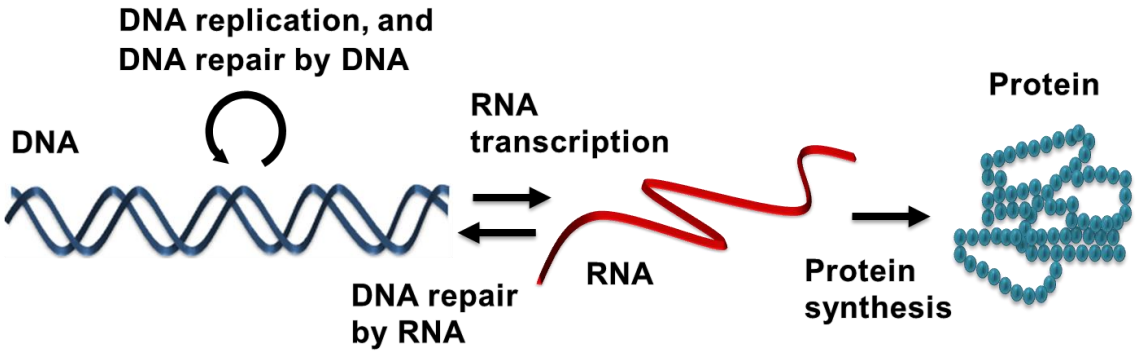


Figure 3.1 Elaboration of the general transfers of genetic information in the central dogma of molecular biology.

DNA is shown in dark blue, RNA in red and protein in turquoise blue. Forward, straight, black arrows represent the direction of the general forms of genetic transfer that occur in cells: DNA → RNA and RNA → protein. The curved, black arrow indicates the general transfer of DNA → DNA occurring in DNA replication and DNA-mediated DNA. The reverse, straight, red arrow marks the proposed more general transfer of RNA → DNA occurring during RNA-mediated DNA repair.

3.4 RNA on the frontlines of DNA damage

3.4.1 RNA-mediated, non-templated DSB Repair

The study of the retrotransposon activity in yeast offered the first glimpse into the effects of RNA in the repair of DNA. Using a HO-endonuclease induced DSB at the mating type locus (*MAT*) in yeast lacking both *HML* and *HMR* donor templates that normally repair *MAT*, Moore and colleagues showed the capture of Ty sequences at the site of the HO DSB. The capture of short ~100bp segments appeared independent of the HR protein Rad52 [55]. Additionally, a *his3* marker gene with an artificial intron inserted into the antisense orientation relative to *HIS3* gene was fused to an inducible Ty1 element on a

plasmid and upon activation of Ty1 transcription and DSB induction, fused *Ty1/HIS3* cDNA was found integrated at the site of an induced DSB [56]. Additional studies also found Ty1 cDNA inserted at sites of DSBs but appeared dependent upon NHEJ factors [37]. Building from these experiments, we illustrate a model of RNA-mediated, non-templated cDNA insertion at sites of a DSB, in which RT activity of Ty elements generates cDNA molecules, that are then hijacked by NHEJ machinery and inserted at sites of DSBs (**Figure 3.2A)**. Ty integrase can also facilitate the integration of non-Ty1 DNA fragments similar to cDNA containing the conserved terminal dinucleotide 5′-TG-CA-3’.[103]

3.4.2 Non-LTR retrotransposon mediated DSB primed reverse transcription

The model for LINE-1 retrotransposition is built from studies of the R2 element in silkworm *Bombyx mori*. It occurs via a target-primed reverse transcription mediated mechanism following endonuclease cleavage [52]. In this model of human LINE-1 retrotransposition, endonuclease activity of the LINE-1 element cleaves DNA, generating a 3’-OH end that is used as a primer for reverse transcription; a second cleavage on the adjacent strand primes the second round of cDNA synthesis and integration of the LINE-1 element into the genome [104]. Studying an endonuclease deficient LINE-1 element in NHEJ-defective Chinese hamster ovary cells, Morrish *et al*. found that LINE-1 retains the ability to retrotranspose, leading to the proposal that LINE-1 elements can integrate at sites of DNA lesions [102, 105]. In support of this model, computational screens identified L1 loci lacking characteristic structural features of target-site primed reverse transcription suggesting an endonuclease-independent L1 insertion at DNA DSBs, providing a role of L1 elements in DSB repair [106]. LINE-1 elements not only transpose themselves but also other retroelements or cellular RNA [100]. In response to DNA damage, LINE-1 reverse

transcriptase can use free DNA ends as a primer for the insertion and repair of DSBs (**Figure 3.2B)**. Could cellular RNA be used in the repair of their own genes with the aid of L1 reverse transcriptase? As described below from experiments in budding yeast, not only the cDNA reverse transcribed from the RNA transcript but also the RNA transcript itself can serve as template for repair of a break in the same DNA gene that generated the transcript RNA. Such RNA-driven DNA repair mechanism may work even more efficiently in mammalian than in yeast cells. In fact, differently from Ty reverse transcription that occurs inside Ty particles in the cytoplasm of yeast cells [107] LINE-1 reverse transcription occurs in the nucleus and can be primed by genomic DNA sequences to which the transcript anneals [102]. Therefore, LINE-1 RT may efficiently synthesize DNA from a transcript template annealed to the broken DNA ends of a mammalian DNA sequence, and thus facilitate DNA repair.

3.4.3 Homologous recombination with cDNA intermediates

cDNA molecules are short dsDNA intermediates of retrotransposition. If cDNA molecules are homologous to genomic regions, could they be used by the HR machinery? Indeed, Ty cDNA of yeast can recombine in a HR manner with homologous Ty elements encoded in the genome, and this process is partially dependent on *RAD52*.[57, 108] DSBs inside of Ty elements can increase recombination between these elements and possibly cDNA intermediates [109]. The ability of Ty elements to generate cDNA molecules of not only Ty RNA but also endogenous transcripts is a proposed model to explain the paucity of introns in the *Saccharomyces cerevisiae* genome and the 5'-bias in genes that retain them [63]. To test this, Derr and colleagues developed an assay using a *HIS3* reporter gene interrupted by an artificial intron inserted in the antisense orientation. The gene cassette

was integrated into an episomal vector under the control of an inducible promoter (galactose inducible promoter, *pGAL1*) in the antisense orientation relative to the native *HIS3* promoter in the sense orientation. Upon galactose induction, transcription and splicing generated an antisense *his3* mRNA which, if reverse transcribed by Ty RT, could be integrated into the genome resulting in His^+ prototrophic cells. The genome contained a deletion of *his3*, which resulted in His^- auxotrophic cells with no homology to *his3* located on the plasmid. Upon galactose induction of the artificial intron containing the antisense *his3* gene cassette, His^+ prototrophic cells were observed at a frequency of around 200 per 10^9, half due to integration into the chromosome and half to integration as a result of *HIS3* cDNA HR with *his3* on the plasmid. The reverse transcription of endogenous cellular RNA and its integration into the genome raises the possibility of recombination events initiated directly by RNA molecules [48]. Using a similar system, in which, a *his3* reporter marker under the regulation of galactose inducible promoter and containing an artificial intron in the antisense orientation was inserted in the yeast genome, Keskin *et al*. detected repair of an induced DSB in *his3* by *HIS3* cDNA via HR. If a cDNA molecule has homology with regions around a DSB, the cDNA can be used as a substrate for repair of that DSBs (**Figure 3.2C)**. Furthermore, inactivation of ribonuclease (RNase) H function by deletion of the RNase H1 gene (*RNH1*) and the catalytic subunit of RNase H2 (*RNH201*) significantly increased the frequency of cDNA recombination, likely through the presence of cDNA-RNA hybrids [40, 90, 101]. *In vitro* and *in vivo* data support the finding that RNase H1 can inhibit reverse transcriptase processivity [110]. However, *in vivo* data suggest that RNase H2 is the preferred RNase H enzyme targeting cDNA in yeast [90]. It is of interest that defects in RNase H2 associated with the inflammatory Aicardi-Goutières

syndrome (AGS) lead to an increase in RNA-DNA hybrids [111]. The yeast AGS orthologous mutations of RNase H2 when combined with RNase H1 null mutation result in elevated cDNA-driven repair of a DSB [90]. This hints at a rise in the levels of RNA-DNA hybrids in AGS patients that are substrates for HR machinery and genome instability. Furthermore, the absence of RNase H activity results in increased RNA-DNA hybrids associated with retrotransposons [112]. It is interesting to speculate if mutants of factors that bind free dsDNA, like those of NHEJ machinery and DSB resection enzymes, could have any effect on cDNA mediate recombination, allowing for increased stability and access to the recombination hardware [113, 114].

3.4.4 Direct RNA-templated DNA repair

As HR is thought to be limited to DNA repair of DSBs in the S/G2 phase of the cell cycle when a sister chromatid is available, a mechanism for accurate repair of a DSB in G0/G1 appears lacking. It was hypothesized that RNA transcripts could provide an alternative template for accurate repair [67]. Studies by Xu *et al.* also proposed the involvement of RNA templated HR to explain the high frequency of homozygosity in rice [70]. To explore the possibility of a direct RNA-templated DNA repair by HR, RNA-only or RNA/DNA chimeric single-stranded oligonucleotides were transformed into yeast cells and used to repair a broken *leu2* marker locus in yeast chromosomal DNA leading to Leu^+ transformants. Deletion of the *SPT3* gene, essential for Ty1 and Ty2 transposition, had no effect on the repair frequency at the *leu2* locus, signifying the ability of RNA-containing oligonucleotides to directly template the repair of a DSB in yeast cells [45]. In addition to yeast, RNA-containing and RNA-only oligonucleotides could also template DSB repair in homologous chromosomal DNA containing the green fluorescent protein (GFP) gene after

DSB induction in human embryonic kidney cells [115]. Building off the work that synthetic RNA oligonucleotides can facilitate the repair of DSBs, Keskin *et al*. showed not only cDNA serving as a template for repair of a DSB through HR but also the direct repair by transcript-RNA in the absence of RNase H function [40]. Transcript-RNA-templated DSB repair appears to be strongly influenced by the proximity of the transcript relative to the site of the DSB. RNA deriving from the same locus (*cis*) as the break has a greater repair frequency opposed to RNA deriving from distant parts of the genome (*trans*) to repair a DSB in a homologous locus [40]. These results demonstrate that actively transcribed DNA experiencing a DSB can repair itself using the genetic information on the transcript RNA. These data also show that the transfer of genetic information from RNA to DNA occurs with an endogenous generic transcript, and is thus a more general phenomenon than anticipated.

Direct RNA-templated DNA repair was revealed in the absence of the genes coding for RNase H [40]. However, direct RNA-templated DNA repair may also be active in the presence of wild-type RNase H function. We know that in wild-type cells DSB repair by transcript RNA is initiated efficiently, but it is quickly suppressed by RNase H1 and H2 activity [40]. Defects in RNases H specifically stimulate DSB repair via HR by RNA-containing substrates in the *cis* system, but do not stimulate DSB repair via HR by homologous DNA donor sequence, or via NHEJ [40, 101]. Thus, it is possible that the inability to detect DSB repair by transcript RNA in wild-type RNase H cells is not due to lack of such a repair mechanism, but rather to limitations of the experimental system adopted. Experiments using synthetic RNA-containing oligonucleotides showed that RNA sequences could directly template DSB repair in yeast chromosomal DNA in wild-type,

spt3 and *rnh1 rnh201* cells with similar frequency [45]. Thus, a system could be developed to allow for a faster DSB repair by RNA, to avoid attack of the template RNA strand by RNase H before completion of repair.

In vivo and *in vitro* experiments suggest that the Rad52 protein can mediate annealing of complementary RNA and DNA strands; however, DSB repair by transcript RNA was not completely abolished in Rad52-null mutants suggesting a Rad52 independent pathway. There are numerous studies which support HR between RNA and DNA molecules. The recombinase Rad51 and its *Escherichia coli* ortholog RecA can promote RNA-DNA hybridization, strand exchange *in vitro,* and R-loop formation *in vivo* [82, 85, 86, 116]. It was found that Rad51 is preferentially recruited to sites of transcriptionally active chromatin experiencing DSBs [117]. More recently, it was suggested that the Cockayne syndrome protein B (CSB) of transcription coupled nucleotide excision repair pathway exhibits the ability to recruit HR factors including RPA1, RAD51C, RAD51 and RAD52 to DSBs sites being actively transcribed in G0/G1. Furthermore, this recruitment is dependent on the presence of transcript RNA [118]. Not only are DSB repair proteins recruited to sites of DSBs in actively transcribed genes, but actively transcribed genes are repaired at a faster rate [119]. Following these recent findings, it is conceivable that transcript RNA may be playing a role in DSB repair at actively transcribed genes. This suggests that transcript RNA can pair with homologous DNA in *cis* preferentially, but also in *trans* at the site of a DSB and facilitate precise end ligation of the DSB ends or extend the 3' end of the DSB by a yet unknown polymerase (**Figure 3.2 D and E**).

We propose a model in which, the transmission of information from RNA to DNA is a more general transfer than commonly accepted and may have a substantial role in

genome stability. We highlight the mechanisms of RNA-driven DNA modification including cDNA-mediated recombination and integration, and direct RNA-templated DNA repair and insertion. Given that RNA molecules are complementary copies of the DNA from which they are derived, it is not an easy task to demonstrate a direct role of RNA in DNA repair. It is likely that we still underestimate the capacity of RNA to mediate non-templated and especially templated modifications into DNA. Thus, new and innovative assays are required to uncover the true roles of RNA in the repair of DNA damage.

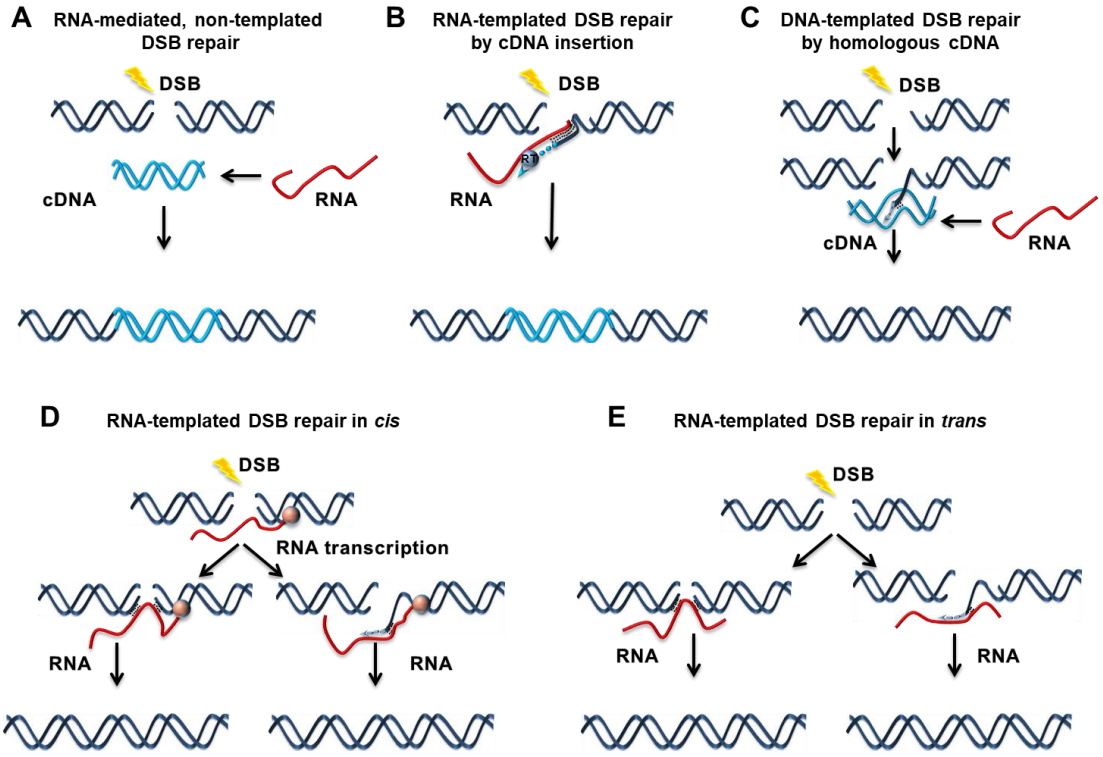


Figure 3.2 Models of RNA-mediated, non-templated and templated DSB repair.

DNA strands are shown in dark blue, RNA in red, and cDNA strands in light blue. DSB induction is pictured by an orange lightning bolt. (A) RNA-mediated, non-templated DSB repair. (B) RNA-templated DSB repair by cDNA insertion. The reverse transcriptase (RT) is represented in blue oval shape. Dotted light blue arrow represents DNA synthesis by the RT. (C) RNA-templated DSB repair by homologous cDNA. The small black line indicates annealing between the DSB end and cDNA. Dotted blue-gray arrow represents DNA synthesis on the template cDNA. (D) RNA-templated DSB repair in cis and (E) in trans. Rose oval shape represents RNA polymerase. The small black line indicates annealing

between the DSB end and the RNA molecule. RNA can bridge the broken DNA ends (left side), or the 3′-DNA end can be extended using RNA as a template (right side). Dotted blue-gray arrow represents DNA synthesis from RNA.

3.5 Acknowledgments

We thank N. Degtyareva and E. Gaucher for critical reading of the manuscript. This study was supported by the National Institute of Health award GM115927 (to F.S.).

CHAPTER 4. FROM "CELLULAR" RNA TO "SMART" RNA: MULTIPLE ROLES OF RNA IN GENOME STABILITY AND BEYOND.

The work in Chapter 4 is part of a larger review article published in *Chemical Reviews* (2018) 118(8):4365-4403. I wrote the manuscript and made the figures.

Michelini F[1], Jalihal AP[2], Francia S[1,3], Meers C[4], Neeb ZT[5], Rossiello F[1], Gioia U[1], Aguado J[1], Jones-Weinert C[1], Luke B[6,7], Biamonti G[3], Nowacki M[5], Storici F[4], Carninci P[8], Walter NG[2], d'Adda di Fagagna F[1,3].

[1]IFOM - The FIRC Institute of Molecular Oncology, Milan, 20139, Italy.

[2]Single Molecule Analysis Group and Center for RNA Biomedicine, Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109-1055, United States.

[3]Istituto di Genetica Molecolare, CNR - Consiglio Nazionale delle Ricerche, Pavia, 27100, Italy.

[4]School of Biological Sciences, Georgia Institute of Technology, Atlanta, Georgia 30332, United States.

[5]Institute of Cell Biology, University of Bern, Baltzerstrasse 4, 3012 Bern, Switzerland.

[6]Institute of Developmental Biology and Neurobiology, Johannes Gutenberg University, 55099 Mainz, Germany.

[7]Institute of Molecular Biology (IMB), 55128 Mainz, Germany.

[8]RIKEN Center for Life Science Technologies, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama City, Kanagawa 230-0045, Japan.

4.1 Abstract

Coding for proteins has been considered the main function of RNA since the "central dogma" of biology was proposed. The discovery of noncoding transcripts shed light on additional roles of RNA, ranging from the support of polypeptide synthesis, to the assembly of subnuclear structures, to gene expression modulation. Cellular RNA has therefore been recognized as a central player in often unanticipated biological processes, including genomic stability. This ever-expanding list of functions inspired us to think of RNA as a "smart" phone, which has replaced the older obsolete "cellular" phone. In this review, we summarize the last two decades of advances in research on the interface between RNA biology and genome stability. We start with an account of the emergence of noncoding RNA, and then we discuss the involvement of RNA in DNA damage signaling and repair, telomere maintenance, and genomic rearrangements. We continue with the depiction of single-molecule RNA detection techniques, and we conclude by illustrating the possibilities of RNA modulation in hopes of creating or improving new therapies. The widespread biological functions of RNA have made this molecule a reoccurring theme in basic and translational research, warranting it the transcendence from classically studied "cellular" RNA to "smart" RNA

4.2 RNA-TEMPLATED DNA REPAIR IN YEAST AND MAMMALS

RNA molecules synthesized during transcription are complementary to the DNA strand that served as their template. Early work demonstrated that RNA could play an indirect role in genome modification and DSB repair if converted into a DNA copy (cDNA) and stitched into damaged sites via NHEJ in yeast and mammalian cells [55, 56, 104, 120].

Not only can these cDNA molecules be inserted in a nonhomologous manner at sites of DSBs, but cDNA can also function as a homologous donor template to accurately repair DSBs via homologous recombination (HR) in budding yeast [40]. However, can an RNA molecule serve directly as a template for repairing/modifying DNA without the need of being converted into cDNA [62, 67]? Indeed, RNA-containing DNA oligonucleotides can serve as templates for gene editing on plasmid or chromosomal DNA in *Escherichia coli*[46, 77, 79]. Similarly, RNA-containing and RNA-only oligonucleotides can serve as RNA donor templates for DSB repair, a phenomenon observed in yeast and human cells[45, 78] In addition, artificial long RNA templates injected in ciliate cells can guide genomic rearrangements[44]. RNA-templated DNA modifications have been proposed to explain the high-frequency non-Mendelian loss of heterozygosity in rice [70]. Moreover, cis- and trans-splicing mechanisms of chromosomal translocation suggest that chimeric RNAs generated by intergenic splicing may play a direct role to guide chromosomal rearrangements [121-125]. A proof of concept that RNA transcripts are recombinogenic and can directly alter the genetic information in chromosomal DNA derives from experiments performed in budding yeast [40]. Given these observations, the importance of RNA-templated repair becomes apparent.

4.3 Molecular details

Keskin *et al.* demonstrated that in *S. cerevisiae* an endogenous transcript can serve as template for repair of a chromosomal DSB in *cis* [40]. The genetic assay was based on the antisense RNA-dependent repair of a nonfunctional histidine auxotrophic marker gene (*his3*). Briefly, an artificial intron (AI) is inserted in reverse orientation relative to *his3*, and antisense transcription is induced (**Figure 4.1)**. While the AI cannot be spliced out of the

sense *his3* transcript, it can be spliced out of antisense transcript. Following the generation of a DSB inside the AI, the pre-existing *his3* antisense transcripts is used as a template for HR, resulting in a functional *HIS3* gene lacking the intronic sequence (**Figure 4.1**). While accurate DSB repair of *his3* is seen in wild-type yeast cells by the formation of histidine prototrophic (His^+) colonies, it is dependent on the reverse transcriptase (RT) activity of yeast retrotransposons, indicating that repair in wild-type cells proceeds through a cDNA intermediate. However, the inability to detect direct RNA-templated DSB repair in wild-type yeast cells may be due to a limitation of the assay used. Indeed, direct RNA-templated DSB repair in wild-type yeast cells is blocked by the function of ribonucleases H (RNase H1 and H2) that cleave the RNA strand of DNA−RNA hybrids. Once the activity of RNase H enzymes is removed, DSB repair is detectable even in the absence of the reverse transcriptase [40]. These results demonstrate the existence of direct RNA-templated DSB repair.

Support for a direct RNA-templated DSB repair mechanism mediated by transcript RNA in *cis* is provided by the dependence on splicing of the antisense RNA. In fact, removal of the 5′- splice site (**Figure 4.1**) eliminates the formation of His^+ colonies. Furthermore, sequencing data and Southern blot analysis support the accurate repair by *cis*-acting RNA, rather than ectopic integration of cDNA transcript from other regions of the yeast genome [40]. Interestingly, even in the absence of the DSB, His^+ colonies are still detectable [40]. This finding suggests that the antisense RNA transcript can even modify DNA without induction of damage, possibly through spontaneous DSBs or nicks in the DNA. Overall, these results demonstrate that RNA can directly transfer genetic information to chromosomal DNA in cis with or without the induction of a DSB, revealing the existence

of a mechanism in which genetic information can flow back from RNA to DNA, beyond the special case of reverse transcription postulated by the "central dogma" of molecular biology [120].

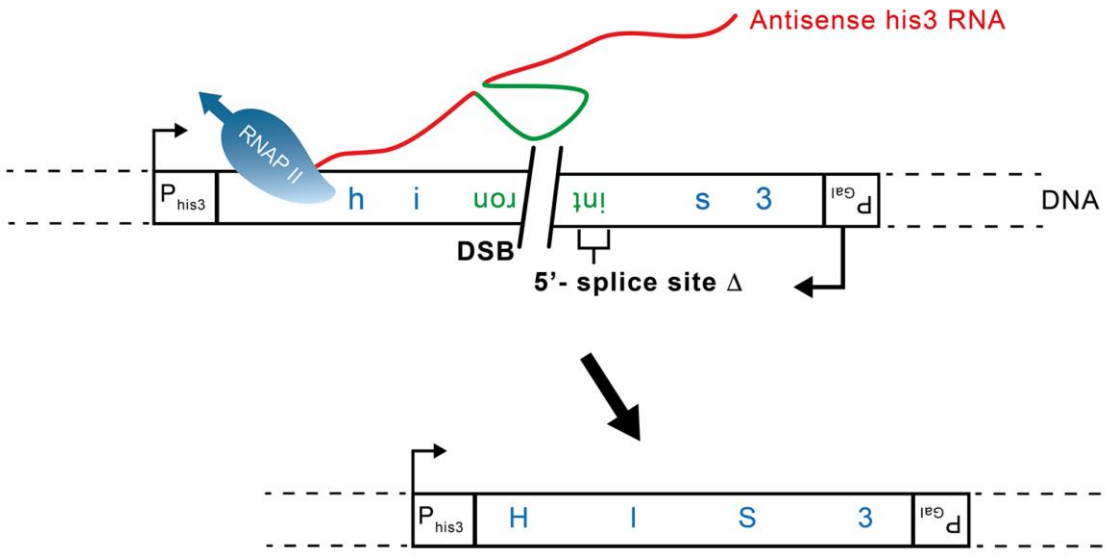


Figure 4.1 Diagram of the system to detect RNA-templated DSB repair.

The *his3* gene is transcribed in the antisense orientation under an inducible promoter and contains an artificial intron that can only be spliced out from the antisense *his3* transcript. Splicing of the antisense RNA and DSB repair by the spliced RNA results in removal of the intron and restoration of a functional *HIS3* gene, which generates His^+ cells. Deletion of the 5′-splice site within the intron sequence is indicated.

4.4 How Does RNA-Templated DSB Repair Work?

Since RNA functions in cis as a donor template in DSB repair of *his3* in the assay described above, the mechanism of DSB repair by RNA is HR. Instead, the sensitivity to RNase H activity indicates that DNA−RNA hybrids must form to transfer information from RNA to DNA. Previous work showed that the RecA recombinase of *Escherichia coli* can promote formation of DNA−RNA hybrids [84, 85]. Yeast RNA-templated DSB repair is strongly dependent on the recombinase Rad52, a fundamental protein in DNA repair by

HR [40, 47]. However, knockout of the *RAD52* gene, while reducing the frequency of DSB repair by RNA by a factor of 10, does not eliminate DSB repair by RNA, indicating that Rad52-independent RNA-templated DSB repair mechanisms do exist. These results in yeast are supported by *in vitro* experiments corroborating the ability of the Rad52 protein to catalyze the annealing of RNA to DNA [40]. Recently, it was shown that purified yeast or human Rad52 protein can catalyze an inverse strand-exchange reaction with DNA or with RNA *in vitro*, a property not observed using the RecA homologue Rad51 recombinase or yeast Rad59, which is important for strand annealing [47]. While RPA inhibits inverse strand exchange between two DNA molecules, it stimulates Rad52-mediated inverse strand exchange between DNA and RNA, possibly via protein−protein interaction with Rad52 [47]. Rad52 also promotes inverse RNA strand exchange with short-tailed or even blunt-ended double-stranded DNA. These results parallel *in vivo* studies demonstrating that RNA-templated DSB repair is stimulated by the overexpression of either yeast or human Rad52 N-terminal domain (NTD) [47]. Rad52 NTD retains the catalytic ability to promote inverse RNA strand exchange but lacks the Rad51 and RPA binding domains [47]. Furthermore, null mutations of the *RAD51* or *RAD59* genes increased the frequency of DSB repair by RNA in yeast [40, 47]. This outcome is thought to occur by curbing the ability of DNA ends to recombine with sister chromatids, funneling repair to an RNA-templated pathway [40]. Moreover, impairment of DNA end processing by defects in *SAE2*, *EXO1*, or *MRE11* genes, which are important for DNA end resection following DSB, either increased or had no change in the frequency of DSB repair by RNA [47]. These data support a model in which Rad52 catalyzes inverse strand exchange between RNA and a nonresected, or little-resected end of DNA at the DSB. RNA then guides break repair by

bridging the broken DNA ends and is used as a template for DNA synthesis to fill the gap, a mechanism that could be mediated by cellular DNA polymerases [40, 47]. If resection is long, RNA-templated DNA repair may require reverse transcriptase for more extensive polymerization.

4.5 Involvement of NHEJ mechanisms

Recently, it has been found that C-NHEJ (classical non-homologous end joining) may play a role in RNA-mediated DSB repair. Following DSBs introduction via bleomycin or ionizing radiation (IR), RNAP II immunoprecipitated with various C-NHEJ and recombination proteins, including LigIV, XRCC4, KU-70, Polμ, DNA-PK, Rad51, and Rad52 [126]. Differently, alt-NHEJ (alternative non-homologous end joining) proteins were absent or far less abundant in RNAP II complexes [126]. The authors reasoned that C-NHEJ proteins may have a role in DSB repair in actively transcribed genes and explored this further. However, in this study, little information is provided on the roles of recombination proteins, which have previously been documented to function at DSBs in active genes [117]. ChIP and quantitative PCR do indeed support the presence of C-NHEJ proteins (53BP1 and LigIV) at sites of DSBs in actively transcribed genes[126]. Importantly, C-NHEJ components were found associated with nascent RNA transcripts by RNA-ChIP and this association significantly decreased following treatment of permeabilized cells with RNase H prior to RNA-ChIP, [126] indicating the formation of DNA−RNA hybrids at DSB sites. This led the authors to suggest that C-NHEJ proteins may aid in an RNA-templated DNA repair mechanism. While RNA-donor oligonucleotides could repair a DSB in human cells in *trans* [46] and an actively transcribed DNA could increase the frequency of end joining ligation of a linearized

plasmid in human cells either directly or via RNA sequences in *trans*, [126] it would be important to determine whether nascent pre-mRNA can template DSB repair in *cis* in mammalian cells. Following up on this possibility, in search of RNA-templated DNA polymerase activity, nuclear extracts of HEK-293 cells strikingly had the capability to copy an RNA template *in vitro*, independently of the major mammalian retrotransposon long interspersed elements (LINE1) [126]. This result highlights the possibility that cellular DNA polymerases may have some RT activity, as shown for yeast replicative polymerases [45], bacterial and archaeal polymerases [127], and some mammalian polymerases [66].

4.6 Models of DSB Repair Mediated by RNA

Overall, these studies unveil an unexpected direct role of RNA in the DSB repair process: RNA may act as a template in repair of DSBs occurring in transcribed DNA [40, 47, 118, 120, 126, 128].160,169,243,259,263,264 An HR model based on the results of experiments in *S. cerevisiae* suggests that a DSB occurring in an actively transcribed gene can be repaired in cis by the transcribed RNA as a bridging template for DNA repair. This process is aided by the inverse strand-exchange activity of Rad52 on dsDNA ends that have limited end resection (**Figure 4.2A**).

In cases of extensive resection, RNA-templated DSB repair could proceed with the aid of an RT. In addition, the RNA transcript can mediate DNA modifications in the absence of Rad52 [40, 128]. Remarkably, RNA retains some ability to modify its DNA gene in *cis* even in the absence of an induced DSB. In this scenario, the RNA partially hybridized to DNA may form an R-loop structure with the intact dsDNA. The failure to remove R-loops from the DNA duplex leads to increase in DNA damage, recombination

rates, mutation frequencies, and loss of heterozygosity[129, 130]. It is generally thought that the majority of R-loop-induced genomic instability stems from encounters between the DNA replication machinery and the altered chromatin environment in the vicinity of an R-loop [131, 132]. If a spontaneous or induced DSB occurs near the R-loop site, repair by C-NHEJ may occur, with the RNA facilitating end ligation by C-NHEJ proteins through end-bridging (**Figure 4B**). Thus, the RNA transcript could be a donor in DSB repair either to allow HR or to guide C-NHEJ, possibly depending on the cell-cycle phase, the types of DSB lesions, and the extent of DNA end resection.

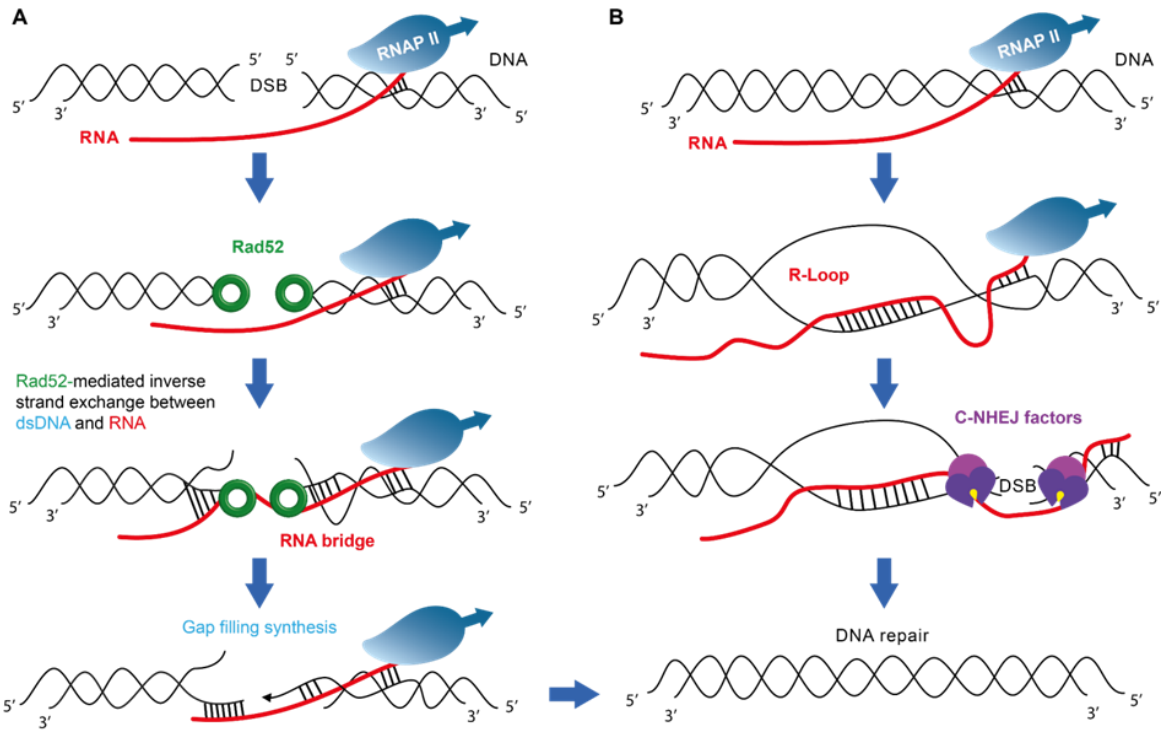


Figure 4.2 Models of DSB repair guided by RNA.

(A) Model of RNA-templated DSB repair via Rad52-mediated inverse RNA strand exchange. The RNA transcript generated from a genomic region that experiences a DNA DSB can anneal with broken DNA ends with the aid of Rad52 in an inverse strand exchange reaction forming an DNA−RNA hybrid that bridges the broken DNA ends and enables transfer of genetic information from RNA to DNA and accurate repair of the DSB. (B) Hypothetical model of RNA-mediated NHEJ repair of DSB. The RNA transcript forms an

R-loop before the DSB occurs. Via a bridging mechanism, likely without the need of a gap-filling step, RNA guides the C-NHEJ protein complex to perform accurate DSB repair

4.7 Acknowledgements

We are thankful to Cinzia Villa for editing figures. F.d'A.d.F.'s lab was supported by the Associazione Italiana per la Ricerca sul Cancro, AIRC (application 12971), Human Frontier Science Program (contract RGP 0014/2012), Cariplo Foundation (grant 2010.0818 and 2014-0812), Marie Curie Initial Training Networks (FP7 PEOPLE 2012 ITN (CodAge)), Fondazione Telethon (GGP12059), Association for International Cancer Research (AICR-Worldwide Cancer Research Rif. N. 14-1331), Progetti di Ricerca di Interesse Nazionale (PRIN) 2010−2011, the Italian Ministry of Education Universities and Research EPIGEN Project, an European Research Council advanced grant (322726). S.F. was supported by Collegio Ghislieri and Fondazione Cariplo (Grant rif. 2014-1215). F.d'A.d.F. and S.F. were supported by AriSLA (project DDRNA and ALS). N.G.W. is supported by National Institutes of Health (NIH) R01 grants GM062357, GM118524, and GM122803, as well as a University of Michigan Comprehensive Cancer Center/Biointerfaces Institute Research Grant. A.J. is supported by the NIH Cellular and Molecular Biology Training Grant T32-GM007315. M.N. is supported by European Research Council Grants 260358 "EPIGENOME" and 681178 "G-EDIT", Swiss National Science Foundation Grants 31003A_146257 and 31003A_166407, and National Center of Competence in Research RNA and Disease. F.S. is supported by the National Institute of General Medical Sciences, NIGMS, of the NIH, grant number GM115927, the National Science Foundation fund with grant number MCB1615335, and the Howard Hughes Medical Institute Faculty Scholar grant 55108574. C.M. is supported by the GAANN fellowship P200A150130-16. G.B. is supported by AIRC grant number 15273. B.L.'s lab

is supported by CancerTelSys (01ZX1302) in the E:med program of the German Federal Ministry of Education and Research (BMBF) and the DFG Heisenberg Program (LU1709/2).

CHAPTER 5. SYSTEMATIC ANALYSIS OF LINKER HISTONE PTM HOTSPOTS REVEALS PHOSPHORYLATION SITES THAT MODULATE HOMOLOGOUS RECOMBINATION AND DSB REPAIR.

The work in this Chapter is part of a research article in *DNA Repair*. I conducted all experiments related to transcript-RNA templated DSB repair in this study.

Kuntal Mukherjee[1], Nolan English[1], Chance Meers[1], Hyojung Kim[1,2], Alex Jonke[1], Francesca Storici[1], Matthew Torres[1]*

[1]School of Biological Sciences, Georgia Institute of Technology; 950 Atlantic Drive NW Atlanta GA 30332

[2]School of Chemistry and Biochemistry, Georgia Institute of Technology; 950 Atlantic Drive NW Atlanta GA 30332

5.1 Abstract

Double strand-breaks (DSBs) of genomic DNA caused by ionizing radiation or mutagenic chemicals are a common source of mutation, recombination, chromosomal aberration, and cell death. Linker histones are DNA packaging proteins with established roles in chromatin compaction, gene transcription, and in homologous recombination (HR)-mediated DNA repair. Using a machine-learning model for functional prioritization of eukaryotic post-translational modifications (PTMs) in combination with genetic and biochemical experiments with the yeast linker histone, Hho1, we discovered that site-specific phosphorylation sites regulate HR and HR-mediated DSB repair. Five total sites were investigated (T10, S65, S141, S173, and S174), ranging from high to low function potential as determined by the model. Of these, we confirmed S173/174 are phosphorylated in yeast by mass spectrometry and found no evidence of phosphorylation at the other sites. Phospho-nullifying mutations at these two sites results in a significant decrease in HR-mediated DSB repair templated either with oligonucleotides or a homologous chromosome, while phospho-mimicing mutations have no effect. S65, corresponding to a mammalian phosphosite that is conserved in yeast, exhibited similar effects. None of the mutations affected base- or nucleotide-excision repair, nor did they disrupt non-homologous end joining or RNA-mediated repair of DSBs when sequence heterology between the break and repair template strands was low. More extensive analysis of the S174 phospho-null mutant revealed that its repression of HR and DSB repair is proportional to the degree of sequence heterology between DSB ends and the HR repair template. Taken together, these data demonstrate the utility of machine learning for the discovery of functional PTM hotspots, reveal linker histone phosphorylation sites

necessary for HR and HR-mediated DSB repair, and provide insight into the context-dependent control of DNA integrity by the yeast linker histone Hho1.

5.2 Introduction

Histones are DNA binding proteins that play an important role in packaging of the eukaryotic genome into chromatin. The basic repeating unit of chromatin compaction is achieved through the association of DNA with an octamer of four distinct histone proteins (H2A, H2B, H3 and H4), around which 147 base pairs of genomic DNA wrap twice to form a nucleosome core particle [133]. Accessibility of DNA to regulatory proteins such as transcription factors is controlled by interactions between non-histone proteins, histones, and DNA. At the hub of these interactions are post-translational modifications (PTMs) commonly found in the N-terminal tails of core histone proteins, which play essential epigenetic regulatory roles in controlling nucleosomal structure and gene transcription [134]. Higher-order states of chromatin compaction arise from the binding of a fifth, linker histone protein (H1), which has been shown to have significant impact on the accessibility of DNA to regulatory and chromatin remodeling factors [135]. Linker histones are comprised of a tripartite domain architecture consisting of a intrinsically disordered N-terminal domain, a globular domain, and an intrinsically disordered and lysine-rich C-terminal domain [136-139]. DNA binding is facilitated through a winged helix (WH) motif within the globular domain that forms asymmetric contacts with the entry or exit ends of a nucleosome as well as through interactions between the C-terminal domain and linker DNA[135, 140, 141].

Yeast *Saccharomyces cerevisiae* harbor a single linker histone, encoded by the gene *HHO1*, which exhibits the same tripartite structure as other H1 proteins but harbors two globular domains (WHD1 and WHD2) instead of one [142-144]. WHD2 extends nearly to the extreme C-terminus of the protein, such that the C-terminal domain is comparatively short (10 amino acids) relative to metazoan H1 proteins, which are long (~100 amino acids). Despite this distinction, yeast and metazoan H1 proteins are different in length by only ~40 amino acids on average – less than half the size of the globular domain. Both WHD1 and WHD2 of Hho1 are structurally similar to the globular domain in human H1 proteins and engage with DNA independently [145, 146].

The cellular function of linker histones has been difficult to study due in part to the existence of multiple variant forms that exhibit seemingly redundant functions. For example, deletion of any one or two of eight histone gene variants in mice has no detectable effect since mice develop normally and there is no change in expression of total H1 levels due to compensatory expression of other H1 variants [147]. In organisms such as *Drosophila melanogaster*, with less H1 variant diversity, recent evidence shows that deletion of H1 leads to de-repression of transposable elements and repetitive DNA sequences, and increases DNA damage due to unregulated R-loop formation [148-150]. Similar results have been found in other organisms with low H1 variant diversity.

Studies characterizing the function of yeast Hho1 have largely focused on presence (*HHO1*), absence (*hho1Δ*) or overexpression of the full-length protein, revealing evidence of both specialized and broad functionality. Early biochemical studies demonstrated that recombinant Hho1 exhibits biochemical characteristics of linker histones and binds stably and stoichiometrically to chromatin *in vitro* [151, 152]. *In vivo,* Hho1 is expressed

substoichiometrically relative to nucleosomal histones (1:37) [152], and deletion of the gene has no impact on cell viability, growth rate, mating, sporulation or basal transcriptional repression of most mRNAs [151, 152]. Deletion of *HHO1* has also been shown not to effect micrococcal nuclease site accessibility *in vivo*, initially suggesting that it does not play a significant role in chromatin compaction [151, 152]. However, more recent evidence suggests that the context under which such observations are made influence the conclusions. For example, evidence from chromatin comet assays and atomic force microscopy measurements suggests that Hho1 maintains normal chromatin structure in wild type cells [153]. In addition, the linker histone has been shown to be essential for chromatin compaction in non-dividing stationary phase cells, where presence of the protein is also anti-correlated with genome-wide gene expression [154]. Finally, longstanding evidence has shown that Hho1 exhibits binding preference for ribosomal DNA (rDNA) [152, 155] where it represses transcription of Pol II genes and is required for maximal Pol I processivity [156]. Hho1 also plays a specialized role in HR. For example, wild type (*HHO1*) cells exhibit suppressed HR at the site of the homothallic switching (HO) double-strand break (DSB), following MMS damage, and in telomeric regions compared to *hho1Δ* cells [157]. Additionally, *hho1Δ* cells exhibit ~3.5-fold greater recombination-dependent loss of a selectable marker from the rDNA locus compared to wild type cells [155]. Considering the several studies wherein presence, absence or overproduction of Hho1 has been used to reveal distinctive effects of the linker histone on DNA biology, it has been suggested that differential expression of Hho1 could serve as a regulatory mechanism. However, recent evidence linking the dynamics of Hho1 protein expression to the dynamics of recombination and repair (as occurs during Meiosis I) have found that little to

no such relationship appears to exist [158, 159]. Consequently, mechanisms that can begin to explain how recombination might be controlled in the presence of Hho1 remain elusive.

We hypothesized that PTM-based regulation of Hho1, which has not been investigated previously, and of eukaryotic linker histones in general, plays an essential role in the maintenance of genome integrity. To approach this problem systematically, we utilized a machine learning model – Systematic Analysis of PTM Hotspots (SAPH-ire) [160-162], to functionally prioritize the compendium of PTMs found on all eukaryotic linker histones. Focusing on phosphorylation, which is readily evaluated through phospho-null and phospho-mimic mutations, we measured the impact on HR as well as other forms of DNA damage repair for 5 observed or conserved yeast Hho1 phosphosites of varying priority score. The results suggest that site-specific H1 phosphorylation plays an important role in modulating HR and DSB repair.

5.3 Materials and methods

5.3.1 SAPH-ire TFx

PTM hotspot analysis was accomplished using the SAPH-ire algorithm as described elsewhere [160, 162], but with a modified neural network model architecture and training scheme [161]. Protein sequence data was retrieved from Uniprot [163] and protein family organization was retrieved from InterPro [164], and PTM data was retrieved from dbPTM [165] and PhosphositePlus (June 2018 release) [166]. Briefly, eukaryotic InterPro family members from IPR005819 corresponding to the linker histone family that harbor at least one experimentally observed PTM are aligned by multiple sequence alignment using MUSCLE [167]. The non-redundant dataset of PTMs [corresponding to Uniprot ID (UID),

native position (NP), residue, and ptm_type] are coalesced into modified alignment positions (MAPs) from which distinct features are extracted and used as inputs for a neural network model trained to classify MAPs as functional or non-functional, resulting in an outcome probability of function potential (SAPH-ire TFx score; see (**Figure 5.1**). SAPH-ire recommendation thresholds are based on pareto front analysis of known functional PTM scores and represent cutoffs for true positive and false positive rates that maximize recall and promote recommendation of ~5% (score >= 0.987) to 10% (score >= 0.950) unknown function MAPs. Validation for the SAPH-ire method has been described previously [160-162]. Both convoluted and deconvoluted data for each MAP can be provided upon request to provide a contextual view of each individual PTM relative to others observed in the same alignment position for the family.

5.3.2 Yeast strains and conditions

A table of yeast strains used or created in this study are provided here (**Table A1**). The yeast *S. cerevisiae* background used predominantly in this study was BY4741 (*MAT*a, *his3Δ1, leu2Δ0, lys2Δ0, ura3Δ0).* Strains containing RNA-templated DNA repair assay are derivatives of YFP17 *(hoΔ hmlΔ::ADE1 matαΔ::hisG hmrΔ::ADE1 ade1 (leu2::pGAL1mhis3AI-ADE3)::HO lys5 trp1::hisG ura3-52 ade3::GAL::HO (his3::HOcs)::Trp1 YCLWTy2-1Δ rnh1Δ::NAT rnh201Δ::hygMX4 pGAL1Δ::klURA3). HHO1-T10A/E, HHO1-S65A/E, HHO1-S141A/E, HHO1S173A/E, and HHO1-S174A/E* mutants were generated using delitto-perfetto as described previously [168], by replacing the native amino-acid to alanine for phospho-null and glutamic acid for phospho-mimic mutants. All mutations were verified by sequencing. For purification and mass

spectrometry, *HHO1* was HA-tagged within the endogenous ORF using delitto-perfetto, resulting in a C-terminal HA epitope in the expressed protein.

5.3.3 Integration of CORE-I-SceI cassette

A table of oligonucleotide sequences used or created in this study are provided here (**Table A.2**). The CORE-I-SceI cassette contains (from 5' end): pGAL1-I-SceI, KanMX4 and KIURA3. In all phospho-null and phospho-mimic strains (MT1013, 1015, 1071, 1072, 1075, 1077, 1078) including WT, a CORE-I-SceI cassette was integrated into the *TRP5* locus. Similarly, in strains CM280, CM1064 (T10A), CM1062 (T10E), CM1042 (S174A) and CM1045 (S174E) a CORE-I-SceI cassette was integrated into the *HIS3* locus. For integration at either locus, chimeric oligo pairs were used consisting of 40 nt homology to the target sequence, 20 nt for the amplification of CORE-I-SceI cassette and 18 nt I-SceI cut site necessary to create DSB [169].

5.3.4 Homologous recombination assay

The HR-mediated DSB repair assay was conducted as previously described [169] Briefly, 1.5 ml overnight cultures were grown in YPD media and the next day split into two separate flasks containing either 50 mL fresh YPD (No DSB) or YP-Lac (DSB) and allowed to grow for 4 hours. YP-Lac cultures were supplemented with 2% galactose to express *GAL1*-I-*Sce*I and induce a DSB. To promote oligo-templated HR-mediated DSB repair, cells grown in either YPD or YP-Lac were washed with water and then transformed with an 80 nt dsDNA repair template (KM167F/R or KM288/289; 1nM), as previously described [170]. Transformed cells were collected and plated with appropriate dilution onto

agar plates containing synthetic complete media lacking either tryptophan (Trp-) or histidine (His-).

5.3.5 PCR-based comparison of I-SceI endonuclease efficiency

Relative efficiency of I-SceI induced DSB was compared across each yeast strain by multi-plexed PCR with KM301/URA3.1 primers to detect the DSB and KM303/KM304 to control for variable genomic DNA input. PCR reactions were carried out on genomic DNA extracted from cells before or after DSB induction.

5.3.6 Non-homologous end joining assay

Yeast cetrometric plasmid YCp50 was cut with NcoI restriction endonuclease at a single site near the middle of the *URA3* gene at 37 °C for 3 hours, resulting in a single linear dsDNA incapable of being repaired through HR in *BY4741*-based strains. Wild type and *hho1* mutant strains were then transformed with the linearized plasmid, plated on media lacking uracil and incubated for 3 days at 30 °C. Surviving colonies were counted and the frequency of Ura^+ colonies calculated. A control strain lacking the essential NHEJ gene *KU70* was included to verify that repair of the plasmid was dependent on NHEJ.

5.3.7 RNA-mediated DNA repair fluctuation assay

Cells were grown in YPlac liquid media overnight, counted and plated in various dilutions to 2% YPGal solid media, grown for 2 days and replica plated to His- media to determine the frequency of repair by RNA.

5.3.8 Construction of diploid cells

Two different methods (M1 & M2) were used to construct diploid cells harboring phosphosite mutations in the *HHO1* locus. In the first approach (M1), mating-type switching in strain BY4741 (*MATa*), YMT1013/YMT1015 (T10A/E), YMT1097/YMT1098 (S65A/E), YMT1071/YMT1072 (S141A/E), YMT1090/YMT1091 (S173A/E), YMT1077/YMT1078 (S174A/E) and YMT1118/YMT1119 (S187A/E) was induced as described previously [171]. Briefly, strains were transformed with plasmid *pGAL1-HO-LEU2,* pre-cultured in SC-Leu media and then inoculated into 50 ml of YP-Lac at a concentration of 10^7 cells/ml and grown overnight. On the following day galactose was added to a final concentration of 2% for 30 min followed by an addition of 1/10 volume of 20% glucose to stop gal induction and growth for 3 more hours. Cells were next diluted and plated onto YPD and grown for 3 days to isolate single colonies. Cells of switched mating type were then crossed with appropriate strains of opposite mating type and diploids selected by single colony growth on agar plates of SC-Trp and SC-Ura. In the second approach (M2), haploid BY4741 strains harboring single point mutations were crossed with BY4742 strains harboring the same point mutations. All replicate results from both methods are reported together.

5.3.9 Cell extracts and immunoblotting

Proteins extracts were generated by glass bead lysis in TCA as previously described [172]. Protein concentration was determined by DC protein assay (Bio-Rad Laboratories), loaded equivalently, and resolved by 12.5% SDS-PAGE followed by immunoblotting with rabbit Hho1 polyclonal antibody (Abcam cat# ab71833) at 1:5000 and glucose-6-phosphate dehydrogenase antibodies (loading control; LC) (Sigma-Aldrich, Cat #A9521) at 1:50,000. HRP-conjugated secondary antibodies (goat-anti-rabbit) were used for

detection of reactant bands by chemiluminescence with ECL reagent (Perkin Elmer Cat #NEL 104001EA). Immunoblots were quantified by high-resolution scanning and pixel densitometry using Image J software [173].

5.3.10 Methyl Methansulfonate (MMS) assay

Yeast strains were grown overnight at 30C in nonselective media (YPD), diluted to a density OD_{600} 0.2, and grown into mid-log phase (OD_{600} 0.8). Cultures were diluted to equivalent density and 5-fold serial dilutions were placed onto YPD plates containing 0.014% MMS followed by outgrowth and imaging.

5.3.11 UV assay

Yeast strains were grown overnight at 30C in nonselective media (YPD), diluted to a density OD_{600} 0.2, and grown into mid-log phase (OD_{600} 0.8). Cultures were diluted to equivalent density and 5-fold serial dilutions were placed onto YPD plates. YPD plates containing diluted cells were then exposed with 7500 μJ UV light in a UV crosslinker (Stratalinker UV Crosslinker) followed by outgrowth and imaging.

5.3.12 Hho1-HA purification and LCMS

Cells from a 1L OD_{600} 0.8 culture of strain YMT1076, in which C-terminally HA-tagged Hho1 is expressed from its endogenous locus and promoter, were lysed with glass beads in 7mL of lysis buffer containing 20mM Tris pH 8.0, 300mM NaCl, 0.1% Triton-X100, 5% glycerol with 0.5mM tris 2-carboxyethyl phosphine hydrochloride (TCEP), 0.5mM PMSF, 50mM β-glycerophosphate, 1x EDTA-free protease inhibitor tablet (Thermofisher), and 250 units of Universal Nuclease (Thermofisher). Hho1-HA was

purified in batch using EZview red anti-HA affinity gel (Sigma #E6779), eluted with excess HA peptide, and resolved by SDS-PAGE (NuPAGE 10% Bis-Tris pre-cast protein gel; ThermoFisher #NP0301). Gels were stained with Coomassie Brilliant Blue (R-250) and bands corresponding to Hho1 were excised and processed for in-gel digestion with Trypsin and endoproteinase Glu-C as described previously [174]. Resultant peptide digests were analyzed by liquid chromatography mass spectrometry (LCMS) with an Ultimate 3000 LC system loaded with an Acclaim Pep Map C18 reverse phase nano-LC column and coupled to a Q-Exactive Plus Orbitrap mass spectrometer (ThermoFisher, Waltham, MA) run in data dependent mode (top 12) with HCD fragmentation. Peptides were separated by gradient LC at 0.3 μL/min using mobile phase A (2% ACN with 0.1% FA) and mobile phase B (80% ACN with 0.1% FA) stepped from 4% B up to 90% B over 150 minutes. MS spectra were acquired from 400-1800 m/z with AGC set to 3E6 and resolution set to 70K. The resulting data were searched against a custom database containing Hho1, Glu-C, and Trypsin sequences using Proteome Discoverer 2.0 and Sequest HT (ThermoFisher). A total of four independent purification/digestion experiments were analyzed in duplicate by LCMS and the MS RAW files of each analyzed together to provide a single consensus report.

5.4 Results

5.4.1 Functional prioritization of eukaryotic linker histone PTMs

To systematically analyze the compendium of PTMs reported for eukaryotic linker histones, we used SAPH-ire, which coalesces 512,015 experimentally observed PTMs across ~38,000 unique eukaryotic proteins organized within ~8,000 protein families and

prioritizes each using neural network models trained to recognize functional PTM sites [160-162]. Within this set, a total of 897 distinct PTMs fall within InterPro classification IPR005819 corresponding to the linker histone (H1) protein family (**Figure 5.1A**). Of these, 10 different PTM types were observed, with phosphorylation, ubiquitination, acetylation, and methylation having the greatest frequency. SAPH-ire coalesces these data into family-specific modified sequence alignment positions (MAPs) of which there were 167 for the linker histone family at the time of this work. Six of these MAPs harbor PTMs identified as functional. Of these, 104 MAPs contain at least one phosphorylation site, 29 of which align with serine or threonine residues in Hho1 (i.e. conserved phosphosites).

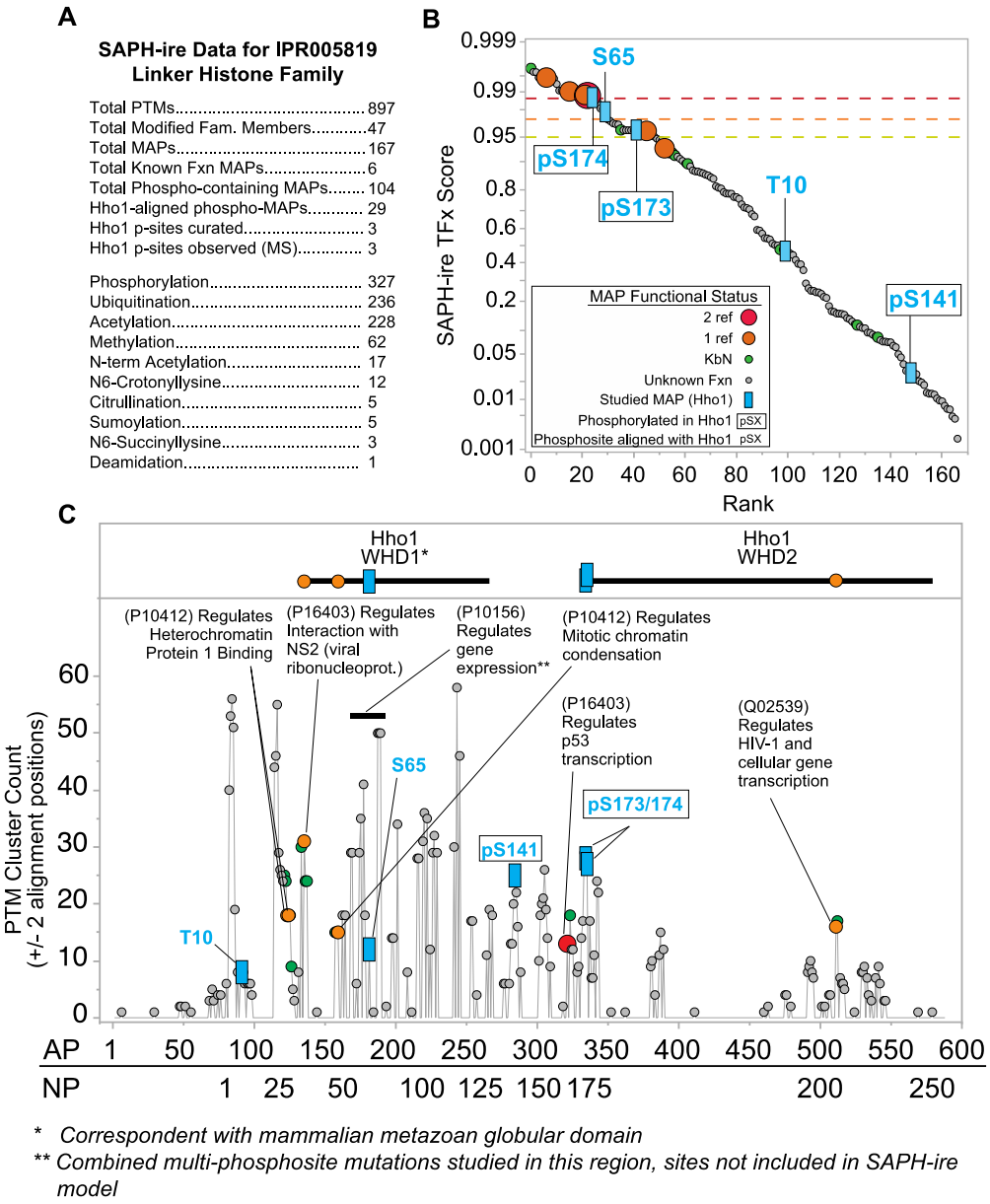
Linker histone MAPs were rank ordered with respect to SAPH-ire score, revealing the enrichment of known-functional MAPs near or above the algorithm's recommendation thresholds at the high priority end of the scale (**Figure 5.1B**). To gain further insight on modification hotspots, we also analyzed PTM clustering across the primary structure of aligned protein family members (see materials and methods), which exposes modification density across the family alignment. Not surprisingly, most PTMs are found clustered within the metazoan globular domain corresponding with Hho1 WHD1, followed by additional clusters in the C-terminal domain that align with Hho1 WHD2 (**Figure 5.1C**). An additional cluster is also evident at the N-terminus of the family (AP 82-86), which is enriched in N-terminal acetylation as well as phosphorylation sites that have not been previously reported for functional impact and some of which are highly ranked by SAPH-ire. Four of the six known functional PTMs localize near the N-terminal end of the metazoan globular domain (AP 135-266) corresponding to regulatory methylation/phosphorylation sites that control protein interaction and DNA binding in

human H1.4 (P10412-K26, S27) or H1.2 (P16403-K34, S36) [175-177], and multiple N-terminal phosphosites (P10256-T35-T54) shown to disrupt gene expression in Tetrahymena when mutated in tandem [178, 179] (see Discussion). Additional functional phosphorylation sites are localized to the C-terminal domain, including human H1.2 (P16403-T146, in MAP-321) near what is the N-terminal end of Hho1 WHD2 that has been shown to regulate p53 transcription [180] and interaction with influenza viral protein NS2 and regulation of interferon-β [176]. Lastly, a phosphorylation site in human H1.1 (Q02539-S183, in MAP-511) has been shown to regulate HIV-1 and cellular gene transcription [181].

In addition to known functional PTMs, PTMs with unknown function are also scored by SAPH-ire – some of which are recommended by the algorithm as likely functional sites of modification (**Figure 5.1B**). Since phosphorylation can be studied genetically using phospho-null and phospho-mimic substitutions, we focused our attention on conserved or directly observed phosphosites in Hho1, three of which (pS141, pS173, and pS174), had been directly observed in the protein previously. The highest ranking of these, pS174, is a proline +1 phosphosite at the N-terminal border of Hho1 WHD2 for which no functional studies have been published and was ranked above the highest SAPH-ire recommendation threshold (0.987) (**Figure 5.1B**) (see Experimental Procedure. Similarly, pS173 was scored above the lowest recommendation threshold (0.950) suggesting its likelihood of functional impact. In contrast, pS141 scored well below the for-recommendation threshold for potential function. Two additional conserved phosphosites, T10 and S65, were also included in the study to evaluate conserved phosphosites that spanned the scoring range and had not been directly observed in Hho1 previously. T10 is near the N-terminal PTM

cluster but received a moderate to low SAPH-ire score, while S65 is a recommended functional phosphosite that falls within the WHD1 domain (**Figure 5.1C**).

Analysis of these conserved phosphosites in yeast could highlight the importance of similarly-positioned phosphosites in humans and other organisms despite not having been observed in yeast directly. The collection of all five investigated phosphosites – both directly observed in yeast Hho1 and conserved in Hho1 – fall within each of the major structural or modification elements in the linker histone family: the N-terminal PTM cluster, the metazoan globular domain (Hho1 WHD1), the C-terminal domain (Hho WHD2), and the inter-domain region between WHD1 and WHD2 (Fig. 1C). These sites also span a broad SAPH-ire score range.



* *Correspondent with mammalian metazoan globular domain*

** *Combined multi-phosphosite mutations studied in this region, sites not included in SAPH-ire model*

Figure 5.1 Functional prioritization of PTMs for linker histone family proteins.

(A) Tabular summary of PTM dataset for linker histone protein family IPR005819. Modified Alignment Position (MAP); Phospho-MAP corresponds to a MAP in which there is at least one observed phosphosite. Linker histone PTMs organized by multiple sequence alignment of family IPR005819 were analyzed using SAPH-ire TFx, resulting in function probability scores (SAPH-ire TFx score) for each MAP in the family (represented by circles) (see *Experimental Procedures*). (B) MAPs plotted by SAPH-ire TFx score and rank order. MAP functional status is indicated by circle size and color, which show the number of published references with evidence supporting functional status of a PTM in the MAP. Green circles correspond to MAPs that are within 2 alignment positions of a MAP harboring a functional PTM (Known by Neighbor; KbN). Blue rectangles correspond to

MAPs that contain yeast Hho1 phosphosites (boxed residue labels) or yeast Hho1 S/T residues that align with phosphosites observed in other family members (unboxed residue labels) that were studied here. SAPH-ire recommendation thresholds are indicated by dashed lines and correspond to: (yellow) score > 0.950 (top 10% recommended); (orange) score > 0.972 (top 7% recommended); (red) score > 0.987 (top 3% recommended). (C) MAPs plotted by PTM cluster count relative to family alignment position (AP) and Hho1 native position (NP). P10412 = human H1.4, P16403 = human H1.2, Q02539 = human H1.1, P10156 = Tetrahymena Hho1 (Native positions 35-54). Note: Tetrahymena p-sites were not curated with other sites used by SAPH-ire and were manually evaluated afterwards.

5.4.2 LCMS analysis of Hho1 phosphorylation sites

To validate previously reported results for Hho1 phosphorylation, and to search for new sites of modification, we analyzed Hho1 by proteolytic digestion and LCMS. HA-tagged Hho1 was purified from log-phase yeast in the presence and absence of induced DSB followed by proteolytic digestion and LCMS (see materials and methods). Nearly 50% coverage of the protein sequence was achieved using a combination of trypsin and Glu-C proteases, which covered the sequence regions surrounding all of the phosphosites in question except for T10. Consistent with previous reports, we observed evidence of phosphorylation at S173 and S174 to a similar degree (15 and 19 PSMs, respectively), but not both simultaneously (**Figure A.1**) [182-185]. We also observed both sites in their non-phosphorylated state (137 PSMs), suggesting that only a fraction of the cellular population of Hho1 is phosphorylated. We did not detect phosphorylation at T10, S65 or S141 in our experiments. Phosphorylation at T10 and S65 has not been reported for Hho1 previously, and our results confirm the absence of phosphorylation at S65, although it cannot be proven outright and may be due to limited sample, digestion or ionization inefficiencies. We conclude that the predominant phosphorylated sites in yeast are S173 and S174 in WHD2 of Hho1.

5.4.3 Hho1 observed and conserved phosphosites recommended by SAPH-ire are required for DSB repair by HR

We next evaluated the hypothesis that Hho1 phosphorylation regulates HR. *Saccharomyces cerevisiae* is widely considered one of the best model organisms in which to study DSB repair and HR due to its genetic tractability and source of longstanding experimental data that serves as reference for new discoveries [186]. Therefore, we used this model to test each of the three phosphosites observed in Hho1 (pS141, pS173, pS174) (**Figure 5.1B and C**; boxed blue labels), as well as the two additional conserved sites (T10, S65) contained in linker histone phospho-MAPs but not known to be phosphorylated in Hho1 (**Figure 5.1B and C**; unboxed blue labels). For each site, a phospho-null substitution to alanine or a phospho-mimic substitution to glutamate was constructed within the endogenous *HHO1* genomic locus of a *BY4741* parent strain engineered to contain a *CORE-ISCEI* construct within the *TRP5* ORF (**Figure 5.2A**). A site-specific DSB was then generated by galactose-induced transcription of *ISCEI*, which produces the ISce-I restriction endonuclease that precisely recognizes and cuts both strands of DNA just upstream from the *CORE-ISCEI* cassette (**Figure 5.2A**), as described previously [169, 187]. To confirm that I-SceI-induced DSB was equivalent in each mutant strain, we conducted a PCR-based test that compares the loss of a PCR product after strand breakage at the site of DSB relative to a control PCR product from a distant locus (**Figure A.2**), finding no evidence of significant bias in cutting efficiency between strains (**Figure 5.2A**). After producing a site-specific DSB in this manner, cells were driven to repair the break via transformation of an 80 bp dsDNA oligonucleotide template that is homologous with *trp5* broken ends that serves as the HR template to reconstitute the functional *TRP5* gene

(see materials and methods), via rounds of Rad52-mediated strand annealing [187]. The advantage of this assay system is that a chromosomal break is induced just before transformation with the template oligonucleotides that contains 40-nt of homologous sequences on either side of the break site, making it very easy to investigate the ability of a DSB to be targeted for recombinational repair.

In the WT *HHO1* yeast strain transformed with the *TRP5* oligonucleotide, DSB-initiated repair of *trp5* (calculated as a frequency of Trp^+ colonies relative to total viable cells) occurs at about 0.3% (**Figure 5.2B**). Deletion of *HHO1* resulted in a modest increase in *trp5* repair. In stark contrast, high priority Hho1 phosphosites (S173 and S174) exhibited near complete suppression of *trp5* repair when substituted with alanine, but behaved like WT Hho1 protein when substituted with glutamate (Fig. 2B). In comparison, phospho-null or mimic mutations to S141 had no significant impact on DSB repair in *trp5*.

We observed similar, although less dramatic effects for the highly ranked S65 (Fig. 2B). This residue aligns with metazoan phosphosites but has not itself been identified as a phosphosite in yeast – suggesting that this position may be an important site of phosphoregulation in humans and other higher eukaryotes in which it occurs. Somewhat surprisingly, phosphosite mutations to T10, a site found in the intrinsically disordered N-terminal region of the family and another site that is not an observed phosphosite in yeast, resulted in an opposite effect to S65, S173, and S174. Indeed, mutants harboring T10E, rather than T10A, were incapable of *TRP5* repair (**Figure 5.2B**). Similar results were observed in the absence of I-SceI DSB induction but in the presence of repair oligos (**Figure 5.2C**), suggesting that Hho1 phosphorylation impacts HR regardless of whether a DSB occurs. Lastly, we observed very little difference in high-dose bleomycin sensitivity

between wild type and Hho1 phosphosite mutant strains, whereas cells lacking the repair protein Rad52 showed complete inability to survive the same conditions (**Figure A.3**). Thus, repair of DSBs that occurred randomly in the genome were not detectably influenced by presence, absence, or phosphosite mutated forms of *HHO1*.

To evaluate the effects of phosphosite mutation on protein stability and cellular abundance, we analyzed the Hho1 steady state protein levels by quantitative western blotting. Results show that none of the mutant proteins exhibit a significant change in protein abundance relative to the WT protein measured with or without induced DSB (**Figure 5.2 D and E**). Taken together, these data suggest that phosphorylation at different positions within H1 linker histones can have dramatically different functional impacts on DSB repair and oligonucleotide-driven HR through a mechanism that is independent of protein stability or abundance but dependent on the genomic context of the DSB.

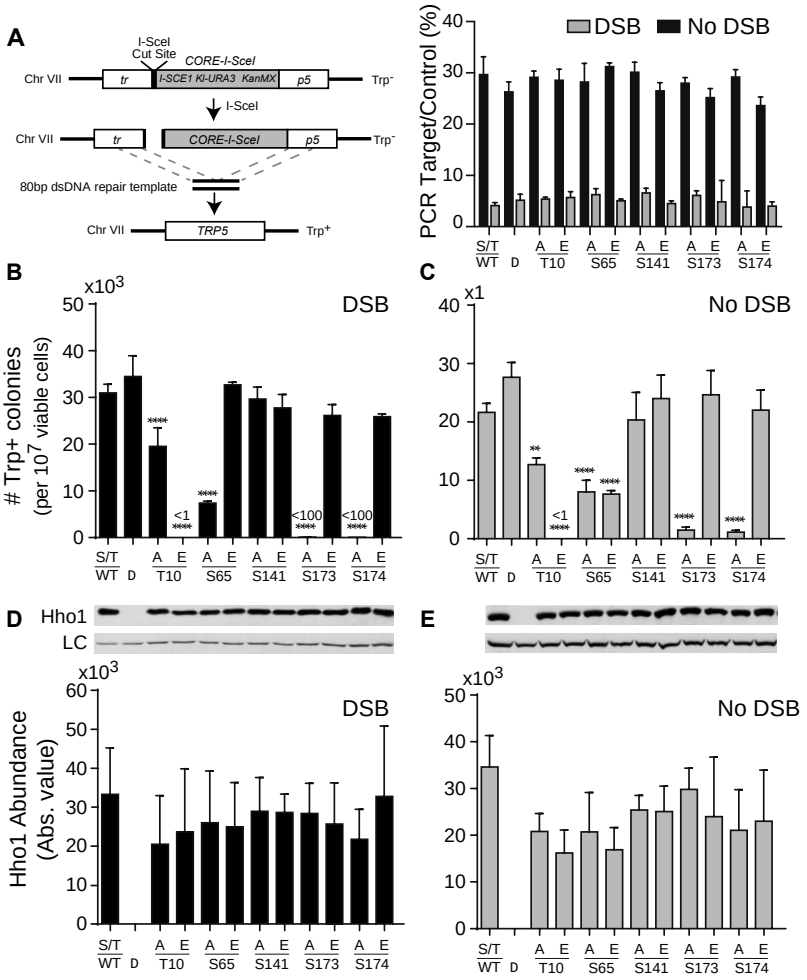


Figure 5.2 Phospho-null and mimic mutations in Hho1 differentially disrupt HR-mediated DSB repair.

(A, left) Schematic diagram of the HR-mediated DSB repair assay. The repair template is an 80 bp dsDNA oligonucleotide transformed into cells after DSB induction. No DSB condition represents cells grown in glucose (i.e. no galactose) and therefore no expression of I-SceI. (A, right) Relative DSB induction between phosphosite mutant strains was quantified by triplicate analysis of multiplexed PCR amplification of the DSB target site (Target) relative to a control site (Control). The ratio between the resulting product in each strain is shown as a percentage in the presence or absence of galactose induction (see also Figure. A2). (B,C) Plot of the number of oligo-mediated Trp+ transformants per 10^7 viable cells resulting from repair of I-SceI-induced DSB for $Hho1^{WT}$ and phosphosite mutants. Note relative response for DSB and No DSB are scaled differently. (D,E) Immunoblot and quantitation of $Hho1^{WT}$ and phosphosite mutant protein abundance in log-phase cells. Error bars represent standard deviation of three or more independent experiments derived from separate yeast colonies. Statistical significance is shown for the comparison between mutant relative to wild type responses and was evaluated using ordinary one-way ANOVA and Tukey's multiple comparisons test.

5.4.4 Hho1 phosphosite mutations do not impact other DNA repair pathways

HR is but one of multiple pathways that cells activate to repair DNA damage. Base excision repair (BER), nucleotide excision repair (NER), and non-homologous end-joining (NHEJ) pathways represent three very common alternative repair mechanisms in eukaryotes [188]. To investigate if Hho1 phosphosite mutations were involved in the BER or NER pathways, we evaluated the ability of each mutant to survive exposure to MMS, a DNA base-alkylating agent that specifically activates the BER pathway *in vivo* [189-191], or UV light, which causes the formation of thymine dimers that activate the NER pathway [192]. None of the Hho1 phosphosite mutants had a major effect on cell survival in either 0.014% MMS or 7500 μjoules UV light (**Figure 5.3A, A.4**). In comparison, control cells lacking *UNG1* and *NTG1*, genes essential for BER, were incapable of surviving MMS treatment, while control cells lacking *RAD1*, an essential component in NER, were incapable of growth under UV stress. We also did not observe any significant difference in the ability of the mutant strains to promote or repress NHEJ in comparison to *ku70*Δ cells, which are completely NHEJ deficient (**Figure 5.3B**). Thus, site-specific mutation of Hho1 at T10, S65, S173, or S174 is essential for efficient HR-mediated DSB repair, but is dispensable for DNA base or nucleotide errors repaired specifically through BER or NER pathways; and for non-templated DSB repair by NHEJ. We conclude that Hho1 phosphosites play a specific role in HR and HR-mediated DSB repair.

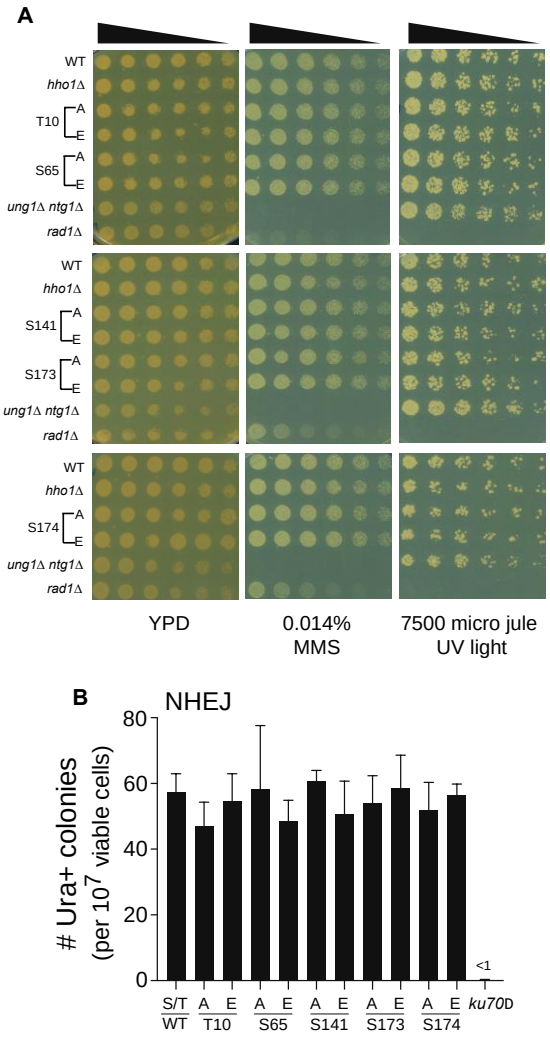


Figure 5.3 Hho1 phosphosite mutations do not impact BER, NER, or NHEJ.

(A) Serial 5-fold dilution of mid log-phase cultured cells harboring wild type or the indicated phosphosite mutants were plated on non-selective media containing 0.014% MMS to induce base excision repair or exposed to UV light to induce nucleotide excision repair, followed by outgrowth. (See **Figure A.3** for replicate tests) (B) Triplicate analysis of wild type and *hho1* phosphosite mutation effects on NHEJ evaluated by plasmid-repair-dependent cell survival on medium lacking uracil. As a control to evaluate the complete loss of NHEJ repair capability, the average response of *ku70*Δ cells is shown in parallel. Error bars represent standard deviation across three independent colony tests.

5.4.5 Hho1 phosphosite mutations regulate DSB repair by a homologous chromosome template in diploid cells

Natural DSB repair through HR relies on extant homology within sister chromatids or homologous chromosomes. To evaluate whether site specific Hho1 phosphorylation is necessary for HR-dependent DSB repair mediated through a homologous chromosome template, we measured repair in diploid yeast without transformation of a synthetic dsDNA oligonucleotide template. In this case, *TRP5* MATa haploid yeast strains harboring individual *hho1* phosphosite mutations were switched to MATα mating type and crossed with *trp5-CORE-ISCEI* MATa cells to produce *TRP5/URA3* diploid cells homozygous for *hho1* and harboring a single *CORE-ISCEI* cassette at one of the *TRP5* loci in Chr-VII (Fig. 4A). Within these strains, DSB repair frequencies are determined by growth on 5FOA and simultaneous sensitivity to G418, indicating removal of the *CORE-ISCEI* cassette by HR repair. As expected, cells expressing Hho1^{S141A} or Hho1^{S141E} displayed no difference in repair frequency, while cells expressing phosphosite mutations at S65, S173, and S174 showed the same relative repair frequencies that we observed in haploid cells transformed with a dsDNA repair template (**Figure 5.4B**). Phosphosite mutations at T10 showed more variable results when compared to the previously observed effect in haploid cells, although trends were generally similar, including those found in the absence of induced DSB (**Figure 5.4C**). Also, in the absence of DSB, mutations at S65 appeared to have a minor effect on HR. These data suggest that Hho1 phosphorylation, particularly at S173 or S174, is important for HR-mediated DSB repair in naturally existing situations where homologous chromosomes rather than exogenous oligonucleotides serve as repair templates.

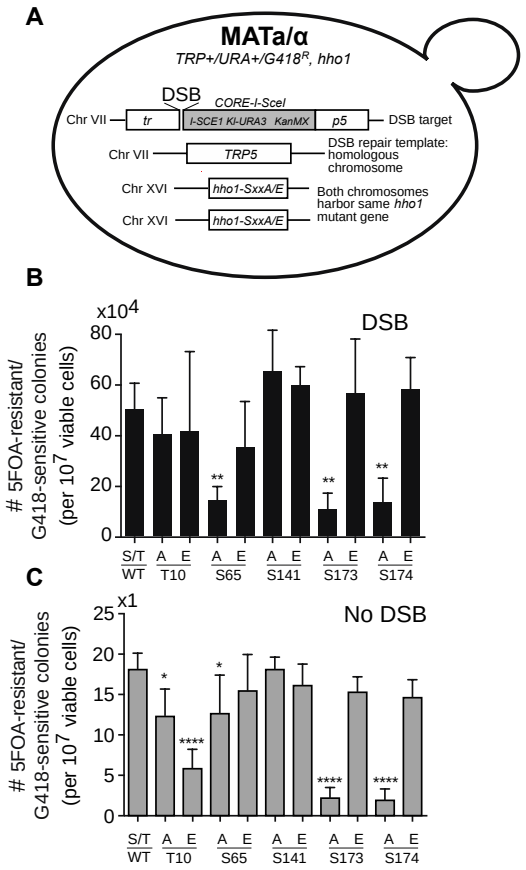


Figure 5.4 HR-mediated DSB repair via a homologous chromosome template is controlled by Hho1 phosphosite mutations.

(A) Schematic diagram illustrating the diploid strain used to evaluate HR-mediated DSB repair through a homologous chromosome template. Repair frequency is measured by the appearance of cells that are both resistant to 5FOA (indicating loss of *KlURA3*) and sensitive to G418 (indicating loss of the *KanMX4* resistance gene). (B) Frequency of 5FOA-resistant and G418-sensitive cells per 10^7 viable cells resulting from repair of I-SceI-induced DSB for $Hho1^{WT}$ and phosphosite mutants in diploid cells. (C) Same as in B, but in the absence of galactose (no DSB). Error bars represent standard deviation of six independent experiments derived from separate yeast isolates for the indicated genotypes. Statistical significance is shown for the comparison between mutant relative to wild type responses and was evaluated using ordinary one-way ANOVA and Tukey's multiple comparisons test.

5.4.6 Impact of Hho1 phosphosite mutations on DSB repair by transcript RNA

Recent evidence has shown that endogenous transcript RNA can be utilized as a HR template for targeted *HO* endonuclease-induced DSB repair in yeast [40]. To

investigate the effect of Hho1 phosphosite mutation on RNA-mediated DSB repair, we expressed Hho1^{S174A} or Hho1^{S174E} from the endogenous *HHO1* locus in the cells used in the Keskin *et al*. study with a slight modification wherein the transcription of the antisense template RNA was controlled by a constitutive promoter (**Meers C. et al., *submitted***). These cells are RNase H-null (*rnh1*Δ *rnh201*Δ), and a DSB is introduced at the HO endonuclease cut site inserted within an inverted intron that disrupts the *his3* marker gene on Chromosome III [40]. DSB repair in these cells can be mediated through either spliced RNA or RNA/cDNA templates resulting from reverse transcription (**Figure 5.5A**) [40]. When RNA was forced to serve as a template for repair, we did observe a small decrease in repair frequency relative to cells expressing wild type Hho1, but no difference between phospho-null or mimic mutations (**Figure 5.5B**). In comparison, we saw no difference in repair between wild type or phosphosite mutants when DNA oligos were provided as repair templates.

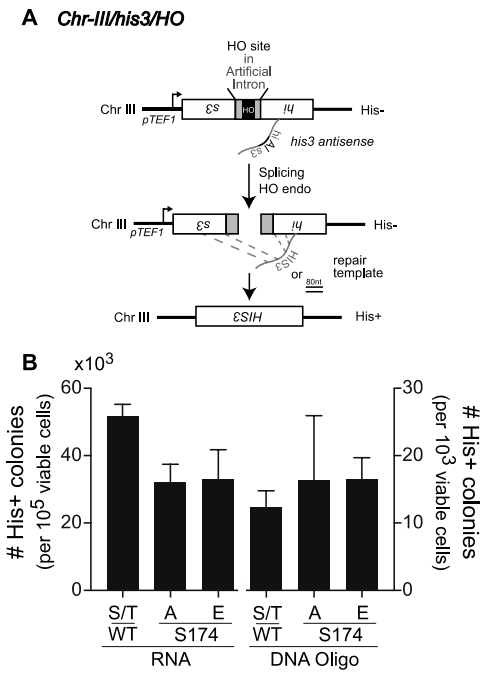


Figure 5.5 RNA/cDNA-templated HR repair of DSBs in wild type and Hho1 phosphosite mutant yeast.

(A) Schematic diagram illustrating the HO-inducible DSB system used to detect DSB repair by transcript RNA or RNA/DNA hybrids. Briefly, an HO endonuclease recognition sequence is embedded within an artificial intron in the middle of the *his3* ORF in chromosome III. Constitutive transcription of the antisense strand produces *his3* antisense RNA that undergoes splicing and reverse transcription to produce a RNA/cDNA hybrid that is utilized as an HR repair template after HO induced DSB. (B) Recombination frequencies measured for cases in which RNA (left) or 80bp DNA oligos (right) serve as the repair template. Where n = 6-12 for the RNA-templated repair assay and n = 4 for the DNA-templated repair assay. Statistical significance is shown for the comparison between mutant relative to wild type responses and was evaluated using ordinary one-way ANOVA and Tukey's multiple comparisons test.

5.4.7 The impact of Hho1 phosphosite mutations on HR-mediated DSB repair is sequence and locus independent

Given the dramatic difference in the repair properties of Hho1 phosphosite mutants in the *Chr-VII/trp5/CORE-ISCEI* versus *Chr-III/his3/HO* assays, we conducted an experiment to control for differences in gene sequence and genetic locus. We began by replacing the *HO* endonuclease recognition sequence of the *Chr-III/his3/HO* strain (**Figure**

5.6A) with the *CORE-ISCEI* cassette, resulting in *Chr-III/his3/CORE-ISCEI* (**Figure 5.6B**). We then compared HR repair in these cells where Hho1^{WT}, Hho1^{S174A} or Hho1^{S174E} were again expressed from the endogenous *HHO1* locus and either HO or I-SceI endonuclease was induced by galactose. Consistent with our initial results, *HIS3* repair in the *HO* strains was unaffected by the presence of Hho1 phosphosite mutants (**Figure 5.6C**). However, in cells harboring Chr-III/*his3*/*CORE-ISCEI* we again observed that repair was significantly repressed in cells expressing Hho1^{S174A} but not Hho1^{S174E} (**Figure 5.6C**). Taken together with earlier results, these data demonstrate that the role of Hho1 phosphorylation in HR-mediated DSB repair is independent of the genetic locus where the DSB occurs.

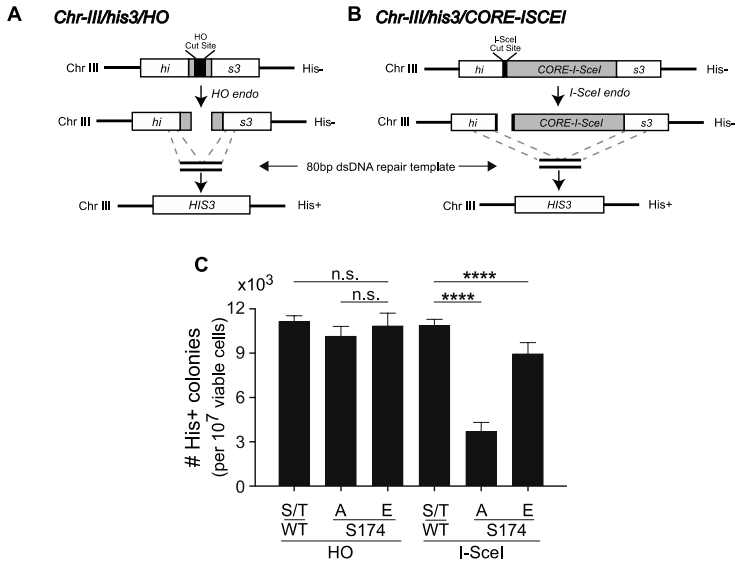


Figure 5.6 DSB repair deficiency of Hho1^{S174A} is independent of DSB genomic location.

(A,B) Schematic diagrams illustrating comparison between *HO* and *CORE-I-SCEI* DSB repair assay constructs, both of which are introduced into the same *his3* locus on chromosome III. (C,D) Plot of the number of oligo-mediated His+ transformants per 10^7 viable cells relative to the time allowed for gal induced DSB by Ho (C) or I-SceI (D). The slope and error of linear trend lines fit to each dataset are indicated next to each curve. Data for C and D were collected from the same experiment but are separated here for clarity. Error bars represent standard deviation of six independent experiments derived from

separate yeast colonies. Statistical significance is shown for S174A vs. WT data and was evaluated using 2-way ANOVA and Tukey's multiple comparisons tests for multiple comparison corrections.

5.4.8 Repression of DSB repair by Hho1^{S174A} depends on the length of DSB-adjacent sequence heterology

Neither gene sequence nor genetic locus can explain the difference in the repair phenotypes of cells expressing Hho1^{S174A}. However, the difference in non-homologous (i.e. heterologous) sequence length at the site of DSB in *HO* and *CORE-ISCEI* strains is considerably large (compare 228bp and 4,800bp), and complete removal of these sequences is required to achieve perfect repair and restoration of a functional selectable marker gene. Therefore, we hypothesized that cells expressing Hho1^{S174A} and forced to undergo oligo-templated DSB repair would be increasingly sensitive to longer lengths of heterologous sequence adjacent to the position of the DSB. To test the hypothesis, we introduced heterologous sequences corresponding to the *K. lactis URA3* gene (1500bp) or *KAN-MX + K.l.URA3* genes (3500bp) immediately downstream of the HO cut site, resulting in a collection of strains with increasing length of heterologous sequence that must be removed to result in accurate repair of *his3* (228, 1729, or 3728 bp), as well as a *Chr-III/his3/CORE-ISCEI* (4800 bp) control (**Figure 5.7A**). We then initiated DSBs by galactose induction of HO endonuclease and compared HR repair in these cells for which Hho1^{WT}, Hho1^{S174A} or Hho1^{S174E} were expressed from the endogenous *HHO1* locus. Consistent with our hypothesis, we found that the relative recombination frequency in cells expressing Hho1^{S174A} decreased significantly with increasing length of embedded heterologous sequence, reaching a baseline equivalent to the *CORE-ISCEI* control (**Figure 5.7B**). In contrast, cells expressing Hho1^{S173E} or Hho1^{WT} exhibited no significant change

in repair efficiency for any of the lengths tested. In the absence of induced DSB, the difference was less detectable except at the greatest heterologous lengths (**Figure 5.7C**). These data suggest that phosphorylation of Hho1 at S174 (and by extension S173 and possibly S65) is likely required for HR-mediated repair of DSBs that occur in cases where extensive heterology exists between the broken DNA strands and unbroken DNA template.

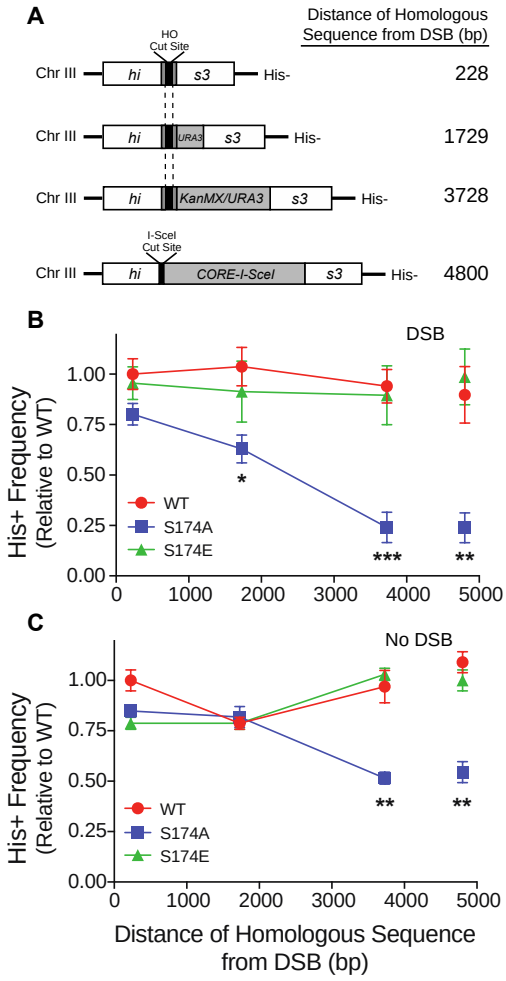


Figure 5.7 Hho1^{S174A} represses HR-mediated repair specifically when long, heterologous DNA is adjacent to the DSB.

(A) Constructs of indicated length were introduced 3-prime adjacent to the *Chr-III/his3/HO* construct shown in Fig. 3A to mimic conditions of variably unaligned break and template strand repair scenarios. Consequently, repair of the HO-mediated DSB requires excision of increasing amounts of heterologous sequence to achieve *his3* repair. (B) Number of His$^+$ colonies per 10^7 viable cells is plotted relative to the average WT response for the 228bp construct in response to DSB. (C) Number of His+ colonies per 10^7 viable cells is plotted

relative to the average WT response for the 228bp construct in the absence of DSB. Error bars represent standard error of the mean for two independent experiments, each containing three independent colony replicates. Statistical significance evaluated using the Holm-Sedak t-test, is shown for S174A vs. WT data.

5.5 Discussion

In budding yeast, which harbor a simplified system with a single linker histone, previous results have shown that the protein controls DNA compaction and HR-dependent DNA repair in a manner that is dependent on deletion or overexpression of the *HHO1* gene. However, under normal cell growth conditions Hho1 expression fluctuates very little, and in cases where it does fluctuate (e.g. Meiosis) such dynamics appear to have no bearing on natural HR processes such as meiotic recombination. Here, we have shown evidence that site-specific phosphorylation of Hho1, independent of protein abundance, plays an essential role in HR and DSB repair by HR.

5.5.1 SAPH-ire contextualizes functional phosphorylation in the linker histone protein family and Hho1

The results shown here provide the first machine-learning-based functional prioritization of PTMs in linker histones and has revealed several interesting features specific to the protein family. While ten unique linker histone PTMs were included in the model, we have focused primarily on phosphorylation, which is the most prominent modification in the family. Site-specific functional phosphorylation of mammalian H1 has not been studied extensively in any organism despite the prevalence of phosphorylation sites found within members of the linker histone protein family (**Figure 5.1A)**. At the time of this report, six independent published articles described a range of phosphorylation-specific effects on H1 proteins. Most phosphosites studied to date occur outside the

globular DNA binding domain of the linker histone (H1.1^{S183}, H1.2^{T146}, H1.4^{S27}). In general, phosphorylation at these sites has been shown (or is hypothesized) to increase DNA accessibility by disrupting H1/DNA interactions [181], or by altering interactions with proteins that repress or enhance gene transcription [175, 176, 180]. One study in human and multiple studies in tetrahymena provide evidence that phosphorylation within the globular domain of human H1.4 (H1.4^{S36}) or similar regions in tetrahymena Hho1 (tetrahymena H1 does not have a canonical globular domain but phosphosites align with regions in the domain family) negatively regulate the association of H1 with chromatin [177, 179, 193]. More recent evidence from systematic CHIP-seq analyses have shown that H1 is significantly depleted from chromatin at sites of HR-repaired DSBs [194]. Thus, the sum of evidence suggests that H1 phosphorylation generally has a net positive effect on processes that require DNA accessibility, and furthermore, that H1 is actively depleted from sites of DSB repaired through homologous recombination.

Our work represents the first functional study of yeast Hho1 phosphorylation, which we have used as a model to test phosphosites spanning the length of the H1 family. To do this, we took advantage of the inherent structure of the SAPH-ire model, which organizes experimental PTM data in the context of protein families through multiple sequence alignment [160-162]. Consequently, we were able to target phosphosites observed in yeast (S141, S173, S174), as well as residues that are conserved with phosphosites observed in other family members (T10, S65). In general, we found that phosphosites recommended for functional study by SAPH-ire (i.e. SAPH-ire TFx score > 0.95; S174, S173, S65) exhibited the most pronounced effects on HR-mediated DSB when mutated, while the most poorly ranked site (S141) had no effect. We also found an

unexpected yet pronounced effect of mutation at T10, which received a mid-level SAPH-ire score below the recommendation threshold. We expect that SAPH-ire functional prioritization data shown here for the H1 family (Tables S1-S5) could prove to be a useful resource for identifying additional functional PTM sites within other family members besides Hho1, including human linker histones.

Consistent with previous studies in human, we find that phosphosite mutation phenotypes in Hho1 depend largely on whether they are inside or outside the globular DNA binding domain, and further, that these sites control the protein through mechanisms not related to its stability. Indeed, phospho-nullifying mutations at S65, S173, or S174 – sites found within WHD1 and WHD2 – is sufficient to repress HR-mediated DSB repair, while phospho-mimic mutations are permissive (**Figure 5.2 and 5.4**). In contrast, functional phosphosites outside of the DNA binding domain (T10) exhibit distinctly opposite DSB repair phenotypes whereby phospho-null mutants are permissive and phospho-mimic mutants are repressive for HR-mediated DSB repair (**Figure 5.2 and 5.4**). This result is consistent with previous data showing that the N- and C-terminal regions of Hho1 exhibit opposite effects on gene transcription from rDNA, wherein analysis of gene truncation mutants has shown that the Hho1 C-terminal WHD2 functions in repression of rDNA transcription while the N-terminal tail and WHD1 derepress transcription [156]. Thus, phosphorylation may be expected to regulate opposite functions when it occurs in the N- or C-terminal regions. Moreover, various forms of indirect regulation have been shown for H1 in regions outside the globular domain that may also play a role. For example, N-terminal phosphorylation of H1.4 (P10412) at S27, in concert with K26 methylation, controls methyl-lysine-specific protein-protein interactions [175]. Despite the fact that the

functional mechanism underlying the T10 phosphosite mutation phenotype is unknown, phosphorylation sites in mammals as well as *C. elegans* are densely clustered within this region near Hho1^{T10} in the linker histone protein family (**Figure 5.1C**), and here we have shown the first evidence that this may be an important regulatory element for Hho1 in yeast, and H1 in general, that is distinct from phosphorylation in the globular domain.

5.5.2 The strong correlation between homology gaps and Hho1^{S174} phosphosite mutants: A window into the contextual nature of Hho1 function?

Longstanding evidence has shown that chromatin compaction is a barrier to DNA damage repair including HR [195, 196]. Like all linker histones, Hho1 can restrict DNA accessibility [144, 153, 154] and cells lacking Hho1 exhibit higher rates of HR [155, 157] Thus, chromatin binding/compaction by Hho1 may decrease DNA accessibility to essential HR machinery, resulting in overall repression of the process. In light of this possibility, we propose that phosphorylation – primarily at S173/174 in WHD2 – may control this feature of Hho1. Our data suggest that phosphorylation occurs on either site, but not both sites simultaneously, and that phospho-null mutation of either site dramatically suppresses oligo-templated (**Figure 5.2**) and homologous chromosome-templated (**Figure 5.4**) HR in the presence or absence of a DSB. We further showed that this effect is independent of the local sequence (*TRP5* or *HIS3*) or chromosomal locus (Chr-VII vs. Chr-III) of the DSB (Fig. 6), but is highly dependent on the presence of heterologous sequence adjacent to the DSB. We have also shown that cells expressing Hho1^{S174A} exhibit an inversely proportional relationship between repair frequency and length of heterologous insert DNA (**Figure 5.7**). Thus, Hho1 appears to have particularly potent impact on HR-mediated DSB repair

between whenever extensive homology gaps exist between DNA break and template strands.

Conversely, we observed several cases in which Hho1 phosphosite mutations have no effect on DNA damage repair, which could provide a possible explanation for their contextual impact. For example, when minimal heterology is engineered between the repair template and the broken strand, Hho1 phosphorylation appears to be largely dispensable (**Figure 5.6 and 5.7**). This is also the case for randomly localized DSBs generated by exposure to bleomycin (**Figure A.3**), and is a likely explanation for why no difference was observed in our RNA repair assay (**Figure 5.5**).

In light of all the evidence, we suspect that Hho1 S174 phosphorylation is involved in regulating either single-strand annealing (SSA) or homology-directed (HDR) HR pathways when extensive heterology is encountered at the break site. Indeed, we show that oligo-mediated repair (**Figure 5.2, 5.5, 5.6 and 5.7**), which occurs exclusively through the SSA pathway [187], and gene conversion in diploid cells (**Figure 5.4**) that occurs through HDR (reviewed in [197]) are both sensitive to *HHO1* phosphosite mutations. Furthermore, both SSA and HDR promote extensive end-resection that would be required to remove long heterologous sequences such as those we have tested in our experiments. This is in stark contrast to the alternative end-joining (ALT-EJ) pathway that requires short homologous sequence lengths and very little resection to accomplish HR [197].

How then does Hho1, and its phosphorylation, participate in the repair process? One possibility is that the Hho1^{S174A} mutant decreases DNA accessibility in a context-dependent manner. For example, Hho1 and S174 phosphorylation may be dispensable

when the distance between the repair template and homologous sequences in the damaged strand is small, but becomes a significant deterrent to repair by SSA or HDR in cases when repair machinery begins to "sense" an expansive gap between these regions (>1,000 bp). Indeed, the substoichiometric abundance of Hho1 relative to nucleosomal histones [151], and its location-specific effect on accessibility and transcriptional silencing at the rDNA locus suggest that the protein, while not utilized for general chromatin compaction, does have the potential to control DNA accessibility [152, 155, 156]. Such control may also depend on the density of Hho1 found at different genomic loci. For example, in rDNA, where Hho1 is densely associated with chromatin, it has significant control over DNA accessibility, transcription, and HR, whilst outside of these regions it appears to have negligible effect [152, 155, 156]. The mechanism by which Hho1 is capable of this is unknown but may involve phosphorylation-based control. Considering the strong dependence of Hho1 phosphosite mutants on heterologous length at the break site suggests a role in resection of long (>1,000 bp) sequence tracts. Such a mechanism may also have implications in controlling recombination of transposable elements (Ty elements in yeast) that are extensively and variably distributed throughout the genomes of different yeast and higher eukaryotes, and could serve as a mechanism to reduce long-distance recombination events or recombination between homologs that contain Ty heterology [198]. Experiments aimed at defining the genomic location and resection dependence of Hho1^{S174A} mutants could shed light on the strong effect of homology gaps that we report here for the first time.

In summary, we have identified a new connection between the site-specific phosphorylation state of Hho1 and DNA repair by HR. We provide genetic and direct evidence that phosphorylation within the yeast-specific globular domain, WHD2, is the

primary phosphorylation-dependent regulator of Hho1. We also provide evidence that phosphorylation in the globular domain (corresponding with S65 in WHD1 of Hho1), while not a directly observed phosphosite in yeast, may have been co-opted for phospho-regulation of H1 in higher eukaryotes. Analysis of these sites has revealed a unique relationship between DSB repair of strands containing extensive homology gaps relative to the repair template that we propose may be linked to locus-dependent differential density of Hho1/chromatin association. Lastly, we provide evidence that the N-terminal tail of linker histones is a functionally impactful and prominent phosphorylation hotspot within the H1 protein family. These represent a few of the many opportunities to better understand linker histones and their complex modes of regulation through post-translational modification.

5.6 Acknowledgements

Funding for the research presented here was provided by NIH NIGMS award R01 GM117400 to M.T. and a subaward from R01 GM118744 to M.T.; and NIH, NIGMS R01 GM115927, NSF, MCB-1615335 and HHMI 55108574 to F.S. Special thanks to Dr. Kirill Lobachev for constructive review of the experiments and manuscript.

5.7 Conflict of Interest Statement

The authors declare that there are no conflicts of interest.

CHAPTER 6. DNA POLYMERASE ζ DRIVES RNA-DNA RECOMBINATION.

The work in this Chapter is part of a research article submitted to *Cell*. I conducted all *in vivo* experiments but those related to Ty-less strains, wrote the manuscript and helped devise experiments.

Chance Meers[1], Havva Keskin[1,2], Gabor Banyai[1], Olga Mazina[3], Taehwan Yang[1], Alli L. Gombolay[1], Efiyenia I. Kaparos[1,4], Gary Newnam[1], Alexander Mazin[3] and Francesca Storici[1]

[1]School of Biological Sciences, Georgia Institute of Technology; 950 Atlantic Drive NW Atlanta GA 30332

[2]Currently at Omega Biotek, Norcross, Georgia 30332, USA.

[3]Department of Biochemistry and Molecular Biology, Drexel University College of Medicine, Philadelphia, Pennsylvania 19102, USA.

[4]Currently at School of Medicine, New York University, New York, New York, USA. 10016, USA.

6.1 Abstract

DNA double-strand breaks (DSBs) are dangerous lesions threatening genomic stability. Fidelity of DSB repair is best achieved by recombination with a homologous template sequence. In yeast, transcript RNA was shown to template DSB repair in the absence of ribonuclease H function. However, molecular pathways of RNA-driven repair processes remain obscure. Utilizing assays of RNA-DNA recombination with and without an induced DSB in yeast DNA, we reveal a major role of translesion DNA polymerase ζ in DSB repair by RNA. Pol ζ catalytic subunit Rev3 with Rev1 and Rad5 is even essential for RNA-templated DNA modification with no DSB. Contrary to DSB repair templated by RNA copied into DNA (cDNA), RNA-templated DNA repair is independent of reverse transcriptase activity of the yeast retrotransposon, and clippase Rad1-10 or Msh2-3. This study characterizes mechanisms of RNA-DNA recombination, uncovering a new role of Pol ζ in transferring genetic information from RNA to DNA.

6.2 Introduction

The preservation of genomic integrity is a delicate balancing act between faithful repair of DNA damage guided by endogenous repair systems and erroneous mishaps of these repair pathways. DNA double-stranded breaks (DSBs) are among the most dangerous types of DNA lesions leading to mutations, chromosome rearrangements and/or inhibiting a cells ability to divide [199]. DSBs are repaired by two main pathways including non-homologous end joining (NHEJ) and recombinational repair. NHEJ proceeds by ligating the broken DNA ends together at the expense of frequent addition or deletion of genetic

information at the DSB site [28]. Homologous recombination involves the exchange of genetic information from a homologous DNA template sequence to the site of DSB [200, 201]. However, genetic information is also transiently deposited in the form of RNA. The transfer of genetic information from RNA to DNA is generally considered to be a specialized process exploited by mobile genetic elements and viruses [202]. We have previously shown that RNA can directly template repair of a DSB in a homologous recombination-dependent (*RAD52*-dependent) manner, when ribonuclease (RNase) H1 and H2 are knocked-out [40, 47, 120, 128]. In addition, mobile genetic elements can reverse transcribe transcript RNA to cDNA-intermediates used in DSB repair [38, 40, 120]. However, little is known about the molecular mechanisms which support RNA-templated DSB repair and DNA modification, and what polymerase enzyme/s can use RNA as template to transfer genetic information to DNA *in vivo*. Here we uncover a major role of translesion DNA polymerase ζ in using RNA as donor in RNA-DNA recombination. Overall, we characterized three mechanisms by which RNA can indirectly or directly template the repair of a DSB, or modify genomic DNA in the absence of an induced DSB. We term these mechanisms cDNA-templated DSB repair (C-TDR), RNA-templated DSB repair (R-TDR), and RNA-templated DNA modification (R-TDM), respectively.

6.3 Materials and methods

6.3.1 Experimental Model and Subject Details

The yeast strains used in this work are listed in (**Table B.1)** and derive from FRO-767 [45]. This strain contains the site-specific homothallic switching endonuclease in the middle of the *LEU2* gene on chromosome III under the galactose inducible promoter

(*pGAL1*). We developed an experiment yeast system consisting of a *his3* gene located on chromosome III containing in artificial intron in the antisense orientation with an homothallic switching endonuclease driven by expression of either *pGAL1* [40] or the constitutive translational elongation factor EF-1α *pTEF* promoter. The native *HIS3* promoter drives transcription from the sense orientation but does not result in functional His3 protein as the artificial intron inserted in the antisense orientation cannot be splicing in the sense orientation. The yeast cells are auxotrophic for histidine (His^{-}) and do not grow on media without histidine. Following galactose induction of the homothallic switching endonuclease and subsequent double-stranded break (DSB) inside of the artificial intron, if the full length antisense *his3* transcript is used for repair of the DSB (in *cis*), this will result in functional *HIS3* gene and protein and growth on media lacking histidine. Accurate repair of functional *HIS3* by ligation of the broken ends at the exon-exon junction via non-homologous end joining (NHEJ) is inefficient in this system (<0.1 out of 10^7 viable cells) (data not shown). Strains CM-278, CM-279, CM-280, CM-281, CM-282, CM-283, CM-284, CM-286 were derived from YS-526, YS- 527, YS-528, YS-529, YS-530, YS-531 as detailed in Table S1, by single-step replacement of the *Kluyveromyces lactis URA3* (*KlURA3*) marker gene with the promoter of *TEF1* EF-1α amplified with short flanking homologies from plasmid p414-TEF-Cas9 [203]. Deletion mutants derived from CM-278, CM-279, CM-280, CM-281, CM-282, CM-283, CM-284, CM-286 were performed by single-step replacement of the opening reading frame of the gene of choice with either the *kanMX4*, *hygMX4*, *natMX4* and/or *KlURA3* unless otherwise indicated and confirmed by PCR. All site-specific modifications or insertions were confirmed by PCR and sequenced. *Saccharomyces paradoxus* strains derive from DG-2204, kindly provided by Dr. David

Garfinkel [204]. HK-692, HK-696 were constructed by the *delitto perfetto* method [205] by inserting a GSKU (GAL1-*I-SceI KanMx4 KlURA3*) cassette with long homology arms to the *LEU2* locus, disrupting *leu2*. The *KanMx4* and *klURA3* markers of the GSKU cassette were then popped out leaving Gal-I-SceI inserted in the *leu2* locus. Deletion mutants derived from HK-692 and HK696 were made as described previously by single-step replacement of the open reading frame of *RNH1* and/or *RNH201*. YCp50pK-Gal-*his3*-AI-*ISce-I* was constructed by PCR amplification of Gal-*his3*-AI-*ISce-I* from genomic DNA (HK-654) with primers adding EcoRI and MluI restriction sites at the ends and subsequent cloning of the EcoRI and MluI-digested PCR product into YCp50pK. The construct was verified by sequencing. YCp50pK-Gal-his3-AI-ISce-I plasmid was used to detect RNA-templated DSB repair events in *S. paradoxus* and *S. cerevisiae* strains in experiments shown in Figure 4B. *S. cerevisiae* FRO-767 strain was used to insert Gal-I-SecI after pop-out of a GSKU cassette at the *leu2* locus, as described above for *S. paradoxus* strains. Strains HK-687 and HK-688 were generated. Successive replacement of *HIS3* ORF with *TRP1* generated strains HK-699 and HK-701, in which all deletion mutants were constructed by replacement of chosen opening reading frames with either the *kanMX4*, *hygMX4* and/or *natMX4* marker gene. Integrated Ty overexpression strains were constructed via the *delitto perfetto* approach. The CORE cassette was inserted into the *CAN1* locus (CM-1099 and CM-1100) and popped out by PCR amplified fragments containing either *pGAL1* from BDG102 plasmid or *pGAL1-Ty1* from pGTyClaI plasmid, which were kindly provided by Dr. David Garfinkel [206], to construct CM-1093, CM-1095, CM-1099 and CM1100, respectively. Deletion mutants derived from these strains were constructed by replacement of the open reading frame of the gene of choice with

either the *kanMX4*, *hygMX4*, *natMX4* and/or *KlURA3*. The *rad50* mutants were created by the *delitto perfetto* method by insertion of a CORE cassette into *rad50* to generate CM-1352, and successive CORE replacement with a PCR product containing the R520H and T853I mutations or the R520H, T853I and D575G mutations to generate CM-1370 and CM-1372 respectively. The constructs were confirmed by sequence analysis. The *rev3* L979F mutant was also created by the *delitto perfetto* method by insertion of a CORE cassette into the *REV3* gene to generate CM-1165 and CM-1166. The CORE cassette was replaced with the sequence of a dsDNA oligonucleotide containing the L979F mutation. The constructs were verified by sequencing. The *pGAL1-REV3* integration strains were constructed by replacement of opening reading frame of *rnh1* with *natMX4* and *rnh201* with *hygMX4* generating CM-1173 and CM-1175. The CORE cassette was then replaced with *pGAL-REV3* and confirmed by PCR and sequencing.

6.3.2 Media preparation

Synthetic dropout, rich YPD (1% yeast extract, 2% peptone, 2% dextrose) and YPGal (1% yeast extract, 2% peptone, 2% galactose) solid and liquid media have been prepared according to standard protocols. Liquid YPLac (1% yeast extract, 2% peptone, 2.7% (v/v) lactic acid)

6.3.3 Fluctuation Assay

Quantitative fluctuation assay in liquid culture was performed to determine DNA repair frequencies as indicated by histidine prototrophic growth. Selected strains were grown in flasks of 50mL YPLac liquid medium for 24hrs at 30 °C. The density of the cultures was determined by counting cells using a hemocytometer and various

concentrations of cells were plated depending on the genotype. In general, per each fluctuation assay, 10^3 cells per sample were plated to YPD solid medium to determine cell survival before DSB induction, grown for 2 days at 30°C and counted. 10^{5-6} cells per sample were plated to YPGal medium to determine survival after DSB induction, grown for 2-3 days at 30°C and counted. In strains lacking *rnh1 rnh201*, 10^{7-8} cells were plated to His^- medium to determine the number of initial His^+ cells before DSB induction, grown for 2 days at 30°C and counted. 10^{7-8} cells were plated to YPGal, grown for 2 days at 30°C and subsequently replica plated to medium lacking histidine, grown for 2-4 days at 30C° and counted to determine the number of His^+ colonies. The frequency of RNA-mediated repair was calculated by dividing the number of His^+ colonies on His^- medium by the number of colonies on YPGal medium and normalized to 10^6 or 10^7 viable cells. The survival was calculated by dividing the number of colonies grown on YPGal medium by the number of cells plated on the same medium.

For experiments using either empty vector (BDG102) or Ty overexpression vector (pGTyClaI), fluctuation assays were performed as described above but cells were grown in 50mL Ura^-Lac medium instead of YPLac medium to maintain plasmid selection. Cells were then plated to Ura^- medium instead of YPD to determine survival, and on Ura^- Gal medium to induce the DSB. 10^7 or 10^8 cells plates on Ura^- Gal were replica plated to His^- medium to determine the repair frequency. The frequency of His^+ colonies was calculated by dividing the number of His^+ colonies on His^- medium by the number of colonies on Ura^- Gal medium and normalizing to 10^7 viable cells. The survival was calculated by dividing the number of colonies on SC-Ura^- Gal medium by the number of cells plated on the same medium.

Experiments using the Ty-less strains were conducted with a plasmid carrying *his3* cassette for the *cis* assay (YCp50pK-Gal-his3-AI-ISce-I). The strains were transformed with YCp50pK-Gal-his3-AI-ISce-I and transformant cells were selected on Ura^- medium. Transformant strains were grown in flasks of 50mL YPLac liquid medium for 24hrs at 30 °C. The density of the cultures was determined by counting cells using a hemocytometer and various concentrations of cells were plated depending on genotype. 10^3 cells were plated to YPD and Ura^- media and grown for 2 days at 30°C to determine survival and plasmid stability. 10^4 cells were plated to YPGal medium and grown for 2 days at 30°C to determine survival frequency following DSB. 10^7 cells were plated to YPGal and grown for 2 days at 30°C and replica plated to His^- medium to determine the frequency of RNA-mediated DSB repair events. The frequency of RNA-mediated repair was calculated by dividing the number of His^+ colonies grown on His^- medium by the number of colonies on YPGal medium and normalizing to 10^7 viable cells. The survival was calculated by dividing the number of colonies on YPGal medium by the number of cells plated on the same medium.

To determine the frequency of RNA-mediated DNA modification events without a DSB, experiments were conducted by either deleting the homothallic switching (*HO*) endonuclease gene or were grown in glucose to repress *HO* transcription. Results from deletion or repression of *HO* were found to be similar. Cells were grown in flasks of 50mL YPLac liquid medium for 24hrs at 30°C. The density of the cultures was determined by counting cells using a hemocytometer and 10^7 cells were plated to His^- medium. The frequency of RNA-mediated repair was calculated by dividing the number of His^+ colonies

grown on His- medium by the number of colonies on YPD medium and normalizing to 10^7 viable cells.

6.3.4 Oligonucleotide and PCR product transformations

Transformation by oligonucleotides (1 nmol) was performed as described [45]. For experiments presented in (**Figures B.1B and B.3D**) oligonucleotides HIS3.F and HIS3.R (**Table B.2**) were used; only oligo HIS3.F for used for experiments presented in (**Figure B.3A**). Induction of the homothallic switching endonuclease DSB was done by incubating cells in 2% galactose medium for 3 h and plating cells to His^- medium to determine repair frequencies. 6.05 µg of PCR product with homology to *HIS3*, generated using primers HIS3.205F and HIS3.205R, were used in transformation experiments presented in (**Figure B.2)**.

6.3.5 Data presentation and statistical analysis

Graphs were made using GraphPad Prism 8.0.2 (Graphpad Software, La Jolla, CA). The results are each expressed as a median and 95% confidence limits (in parentheses), or alternatively mean with range (in parentheses) when indicated. Statistically significant differences between the His^+ frequencies were calculated using the nonparametric two-tailed Mann-Whitney U test [207]. Significance of comparisons is indicated as *, $P < 0.05$; **, $P < 0.01$, and ***, $P < 0.001$.

6.3.6 Western Blot Analysis

Whole cell protein extracts were isolated by collecting $5 \times 10^8 – 1 \times 10^9$ cells per sample. The *S. paradoxus* and *S. cerevisiae* cultures were grown in YPD media to OD600

0.5, whereas for overexpression experiments containing *S. cerevisiae* cells expressing Ty under the *pGAL1* promoter cells were grown in YPLac medium overnight to OD600 0.3. Then, galactose at 2% (v/w) final concentration was added to the medium and cells were shacked in the incubator at 30 °C for 6 hours [40]. After harvesting by centrifugation, cells were washed with PBS (Corning, 21-040-CV) and kept at -80 °C until processing. Cells were resuspended in lysis buffer [25 mM TRIS pH 7.5, 100 mM NaCl, 1 mM DTT, 10 mM EDTA, 1x protease inhibitor cocktail (Roche, 11836170001) 0.05% NP-40 (Thermo, 28324)] on ice and ~100 µl glass beads were added before bead beating using the Genie disruptor machine (1 minute shaking, 1 minute rest, 5 times in cold room). The supernatant was carefully removed, centrifuged and the concentrations determined by Bradford reagent (Biorad, 500-0006) and BSA standards (Biorad, 500-0207) and a spectrophotometer. Successively, 5 µg protein extract was loaded on a 12 % TRIS-glycine SDS PAGE gel. After running, the gel was blotted on an Amersham nitrocellulose membrane (GE, 10600003) overnight at 4 °C. The membrane was blocked with 5 % milk and tris buffered saline with tween for 1h at room temperature, then either Actin (ab170325) or B8 antibodies [208] were added in 1:3000 dilution in 5 % milk + TBST and incubated overnight at 4 °C. The membrane was washed 3 times for 10 minutes with TBST and respective secondary antibodies (Thermo, 31460, 31430) were added in 1:5000 dilution in 5 % milk + TBST and incubated at room temperature for 1h. The membranes were washed again 3 times and developed using ECL Western Blotting solution (Thermo, 32106) according to the manufacturer's instructions.

6.3.7 RNA isolation and Gene Expression measurements

Total RNA was isolated using the Hot Phenol method [209]. Approximately 3 x 10^8 – 5 x 10^8 cells were collected by centrifugation and washed once with ice cold PBS and stored at -80 TBST °C overnight. Cells were then resuspended in 500 µl TES solution (10 mM TRIS-HCl, pH 7.5, 10 mM EDTA, 0.5% SDS) and 500 µl acid phenol (VWR, 0981.400ML) was added. Samples were vortexed vigorously and incubated at 65 °C for 1 hour with brief vortexing every 15 minutes. Cells were chilled on ice for 5 minutes, then centrifuged for 5 minutes at full speed on a tabletop centrifuge. The upper aqueous supernatants were transferred to clean RNase free tubes (Fisher, AM12450) and extracted twice by 500 µl chloroform. Total RNA was precipitated at -80 °C by sodium acetate (3M, 1/10 volume) and ethanol (2.5 volume) and washed by pre-chilled 70 % ethanol. RNA pellets were resuspended in RNase free water and purified using RNeasy mini kit (Qiagen, 74104) according to the manufacturer's instructions. This was followed by DNase treatment using the Turbo DNA-free kit (Fisher, AM1907) according to the manufacturer's instructions. RNA concentrations were determined using a Nanodrop 1000 machine and 1 µg of total RNA was used to reverse transcribe to cDNA using the iScript cDNA synthesis kit (Biorad, 1708891) according to the manufacturer's instructions. The final product was diluted 10 times to 200 µl and 2 µl of this product was used as template for each qPCR reaction. Quantitative real time measurements were performed according to the manufacturer's recommendations using the 96-well Step One Plus Real Time PCR system (Fisher, 4376600), SYBR select master mix (Fisher, 4472918) and the primers HIS3Q.1, HIS3Q.3 for *his3* and ACT1Q.F, ACT1Q.R for *ACT1* shown in Table S2. Relative fold changes were determined by normalizing to the actin levels of non-induced *pTEF* system samples for each background. At least two duplicate measurements were performed for at

least two biological repeats and error bars show the standard deviation for each sample [210].

6.3.8 In-vitro D-loop and R-loop assay

ScRad52 (450 nM) was incubated with a 48-mer 32P-labeled ssDNA (no. 211; 3 μM, nt) or a 48-mer RNA (no. 501; 3 μM, nt) of identical sequence in buffer containing 25 mM Tris·acetate (pH 7.5), 2 mM DTT, 0.2 mM magnesium acetate, 20 mM KCl (added with the protein stock) and 100 μg/ml BSA for 15 min at 37 °C. D-loop formation was initiated by addition of supercoiled pUC19 dsDNA (67.2 μM, nt). Aliquots (10 μl) were withdrawn at indicated time points and deproteinized by incubation in 1 % SDS, 1.6 mg/ml proteinase K, 6 % glycerol and 0.01 % bromophenol blue for 15 min at 37 °C. Samples were analyzed by electrophoresis in 1 % agarose-TAE (40 mM Tris·acetate, pH 8.0 and 1 mM EDTA) gels. The gels were dried on Amersham Hybond-N+ membranes, and then visualized and quantified using a Typhoon FLA 7000 Phosphor Imager (GE Healthcare). The yield was expressed as a percentage of the total plasmid DNA.

6.4 Results

6.4.1 Constitutive expression of the RNA donor promotes DSB repair by template RNA in cis

To better characterize the mechanism regulating RNA-templated DSB repair (R-TDR), we modified the assay that we previously developed to study DSB repair by transcript RNA in *cis* in yeast cells [128]. To drive transcription of the non-coding antisense *his3* RNA, which serves as homologous template for DSB repair, we replaced the galactose

inducible promoter *pGAL1* with the constitutive translation elongation factor EF-1α promoter (*pTEF*) (**Figure 6.1**) to make strains CM-278, 279 (**Table B.1**). In this way, the template RNA is already actively transcribed at the time of DSB induction. In brief, the genetic assay comprises a *his3* gene interrupted by an artificial intron in the antisense orientation containing a homothallic switching (HO) endonuclease site, which is driven by either the inducible *pGAL1* or the constitutive *pTEF* promoter from the antisense orientation (**Figure 6.1A**). Galactose induction of the HO endonuclease gene expressed under a *pGAL1* promoter and located at another genomic locus, results in a DSB inside the artificial intron within the *his3* gene (**Figure 6.1A**). If the spliced antisense transcript RNA driven by either the *pGAL1* or *pTEF* promoter is used to template repair of the DSB by removing the artificial intron from its own *his3* DNA (in *cis*), a functional *HIS3* gene is formed producing His^+ colonies.

In the *cis* DSB repair assay, wild-type RNase H cells repair the *his3* DSB by converting the spliced antisense transcript into a complementary DNA (cDNA) donor for homologous recombination. The cDNA-templated DSB repair (C-TDR) process was shown to depend on the expression of the donor transcript and the activity of the yeast retrotransposon Ty, because deletion of the antisense promoter or the *SPT3* gene abolished C-TDR [40] (**Figure 6.1B**). When we expressed the *his3* antisense donor from the constitutive promoter, *pTEF*, we found that the frequency of His^+ colonies decreased by a factor of ten (**Figure 6.1B**). This is likely due to lower expression level of *his3* RNA from *pTEF vs PGAL1* (**Figure B.1A**). RNase H1 and H2-null cells (*rnh1 rnh201*) in the *pTEF* system show an increase by a factor of over one hundred-fold in the frequency of His^+ colonies. This increase is dramatically reduced by deletion of the promoter [40], but not by

deletion of *SPT3* **(Figure 6.1B)**. A high frequency of His^+ colonies is observed in *rnh1 rnh201 spt3* cells of the *pTEF* system, indicating that this *HIS3* repair occurs by R-TDR. Differently from what is observed in wild-type RNase H cells, we found that the frequency of His^+ colonies did not diminish, but even increased (over a factor of three) when the antisense RNA was driven by *pTEF vs PGAL1* in *rnh1 rnh201 spt3* cells (**Figure 6.1B**). This opposing trend of His^+ frequencies obtained when the antisense *his3* RNA is expressed from *pTEF* vs *pGAL1* is not attributed to increased level of antisense *his3* RNA in *rnh1 rnh201 spt3* cells because steady-state levels of *his3* remain higher with *pGAL1 vs pTEF* in these mutant cells (**Figure B.1A**). Transformation with DNA oligonucleotides (HIS3.F and HIS3.R, **Table B.2**) designed to repair the DSB in the *his3* gene showed slight (less than 2-fold) increase of repair in the *pTEF* vs *pGAL1* system (*rnh1 rnh201 spt3*) suggesting a possible increase in DNA breakage in the *pTEF* system (**Figure B.1B**). These results suggest that continuous transcription of the antisense donor RNA from the constitutive promoter facilitates recombination of the broken DNA ends with the RNA template in *cis*, increasing the frequency of R-TDR. High induction of transcription on the other hand stimulates production of cDNA-intermediates to repair DSBs via C-TDR.

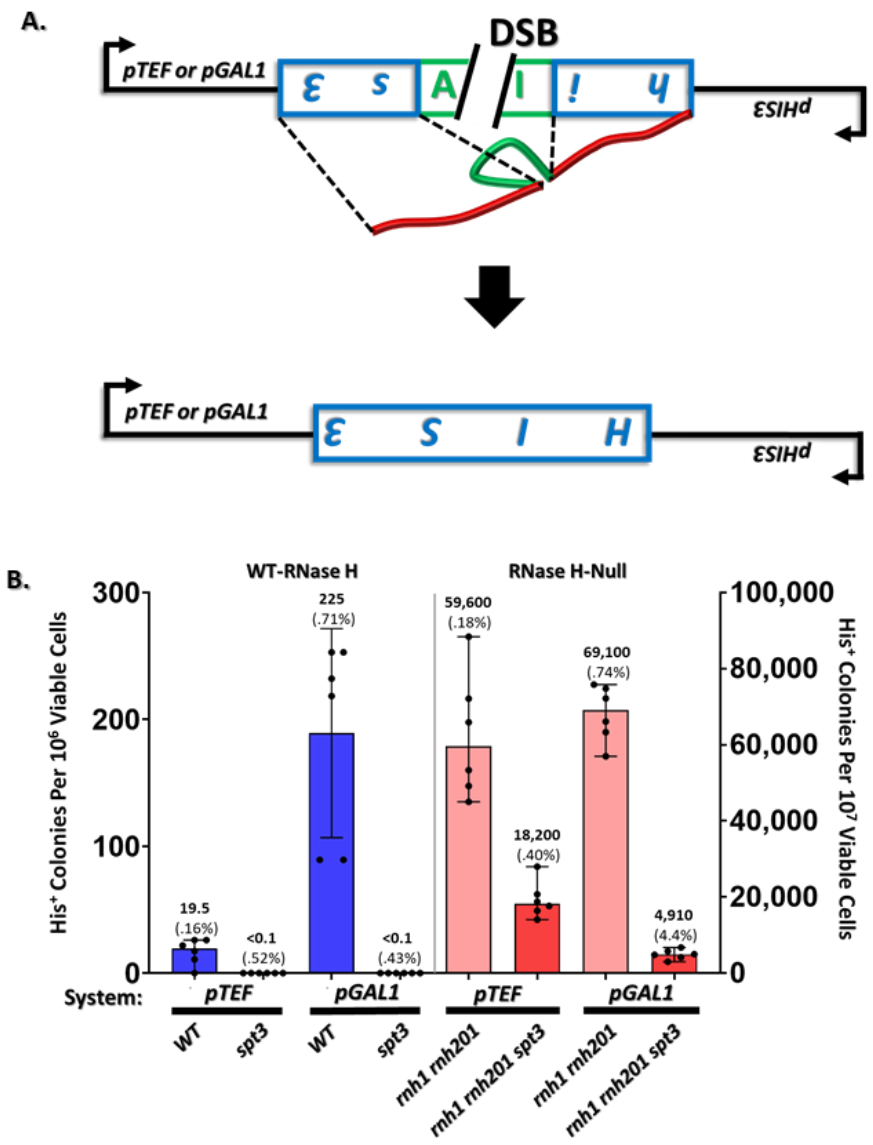


Figure 6.1 DSB Repair by RNA in *cis* Is Facilitated by Constitutive Expression of the Template RNA.

(A) Scheme of the genetic system used to detect R-TDR in yeast cells. The system contains a *his3* gene interrupted by an artificial intron (AI, in green) in the antisense orientation under either the galactose inducible promoter (*pGAL1*) or the constitutive translation elongation factor promoter (*pTEF*). Following transcription and splicing in the antisense orientation, the antisense transcript-RNA (shown in red) is used to guide removal of intronic sequence coded in DNA following a DSB inside the intronic sequence. This results in a functional *HIS3* gene and yeast cell growth on medium lacking histidine. The DSB is generated by a galactose inducible HO endonuclease present on chromosome III. (B) Fluctuation assay showing frequency of His^+ colonies per 10^6 or 10^7 viable cells following DSB induction in wild-type RNase H or RNase H-null cells. The donor antisense RNA is

either expressed from the constitutive *pTEF* or the inducible *pGAL1* promoter. Cell genotypes are indicated. Individual frequencies are plotted. Bars represent median with 95% confidence interval. The median is shown above each bar and survival following DSB induction is shown in parentheses; N=6. *P*-value are shown in (**Table B.3**).

6.4.2 R-TDR requires Rad52 and is independent of NHEJ proteins

Previous results using the inducible antisense *his3* RNA expressed under *pGAL1* showed that R-TDR requires the recombination enzyme Rad52 for an inverse strand exchange reaction. In contrast, the frequency of R-TDR increases in *rad51*-null or *rad59*-null cells, likely because these mutants suppress competition for DSB repair by the intact sister chromatid [40, 47]. Here, using the constitutive *his3* RNA expressed under *pTEF*, we observed a factor of 100-fold decrease in the frequency of R-TDR in the absence of *RAD52*, while knock out of *RAD51* or *RAD59* showed no effect. Differently, C-TDR requires all these recombination proteins in the constitutive system (**Figure 6.2A**).

Recent studies have suggested a possible role of RNA facilitating non-homologous end joining (NHEJ) [126]. In the *cis* assay with *pTEF* (**Figure 6.1A**), we found that elimination of NHEJ components (*KU70* or *DNL4*) markedly enhanced the frequency of His^+ colonies but dramatically decreased cell viability following DSB induction (**Figure 6.2B**). We also observed similar effects with the loss of the NHEJ DNA polymerase 4 (*POL4*) (see **Figure 6.5A**). *RAD50* mutants (R520H T853I and R520H T853I D575G) identified in a random mutagenesis screen also showed increased frequency of R-TDR **(Figure 6.2C)**, in line with inhibition of the NHEJ pathway. These results suggest that while it is possible that RNA facilitates NHEJ, NHEJ enzymes are not needed for R-TDR. This confirms that transcript RNA can work as a template for DSB repair in *cis* via a homologous recombination mechanism.

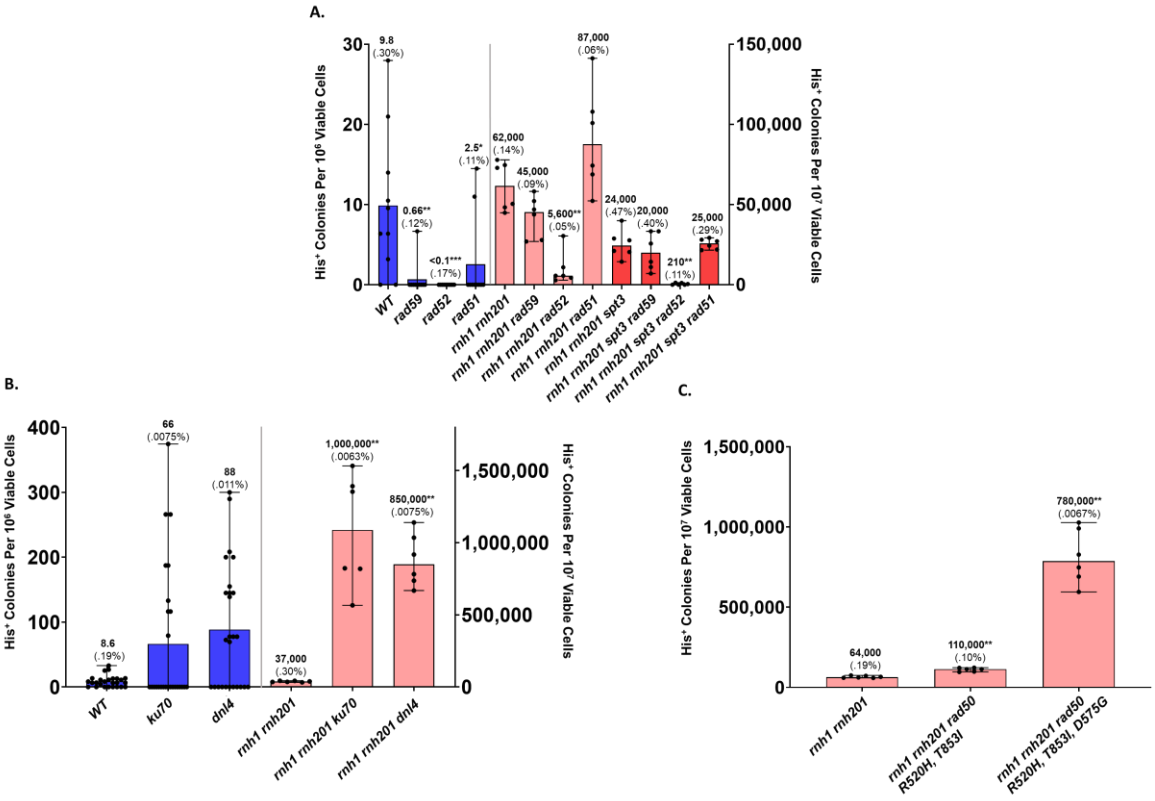


Figure 6.2 R-TDR requires Rad52 but not NHEJ proteins.

(A) Fluctuation assay showing frequency of His^+ colonies per 10^6 or 10^7 viable cells following DSB induction in different recombination mutants. Individual frequencies are plotted. Bars represent mean with range (WT, *rad59*, *rad52*, *rad51*) or median with 95% confidence interval (*rnh1 rnh201, rnh1 rnh201 rad59, rnh1 rnh201 rad52, rnh1 rnh201 rad51, rnh1 rnh201 spt3, rnh1 rnh201 spt3 rad59, rnh1 rnh201 spt3 rad52, rnh1 rnh201 spt3 rad51*). The genotype of the samples is indicated under each bar. The median or mean of each genotype described above is shown above each bar and survival following DSB is shown in parentheses; N=6-10. *P*-values are shown in (**Table B.4)**. (B) Fluctuation assay showing frequency of His^+ colonies per 10^6 or 10^7 viable cells following DSB induction in NHEJ mutants. Individual frequencies are plotted. Bars represent mean with range (WT, *ku70*, *dnl4) or* median with 95% confidence interval (*rnh1 rnh201, rnh1 rnh201 ku70, rnh1 rnh201 dnl4*). The genotype of the samples is indicated under each bar. The median or mean of each genotype described above is shown above each bar and survival following DSB is shown in parentheses; N=6-26. *P*-values are shown in (**Table B.5**). (C) Fluctuation assay showing frequency of His^+ colonies per 10^7 viable cells following DSB induction in *rad50* mutants. Individual frequencies are plotted. Bars represent median with 95% confidence interval. The genotype of the samples is indicated under each bar. The median is shown above each bar and survival following DSB is shown in parentheses; N=6. *P*-values *rnh1 rnh201* - *rnh1 rnh201 rad50* R520H, T853I (0.0022**), *rnh1 rnh201* - *rnh1 rnh201 rad50* R520H, T853I D575G (0.0022**).

6.4.3 3' non-homologous tail removal by Rad1-10 and Msh2-3 is dispensable for R-TDR

To further characterize molecular components of R-TDR, we investigated the role of DNA clippases, which are important players in DSB repair by homologous recombination [211, 212]. We discovered that loss of *RAD1-10* strongly decreases the frequency of C-TDR but not R-TDR. In wild-type RNase H cells, the frequency of His^+ colonies dramatically decreased upon knockout of *RAD1* or *RAD10*, while in *rnh1 rnh201 spt3* cells, in which repair is directly templated by transcript-RNA (R-TDR), knock out of *RAD1-10* shows only minor reduction in His^+ frequency. This suggests that *RAD1-10* clippase is important for C-TDR but has only have a minor impact on R-TDR (**Figure 6.3A**). Therefore, loss of either *RAD1-10* or *SPT3* [40] allows separation of R-TDR from the C-TDR repair mechanism, as both inhibit C-TDR. Similar to DSB repair by *his3* cDNA, DSB repair by a *his3* linear DNA molecule generated by PCR was strongly reduced in *rad1*-null cells (**Figure B.2**). These results show that *RAD1-10* is important for DSB repair by a homologous DNA template provided exogenously, likely to remove 3' tails of cDNA or PCR product that are not used in the recombinational repair process [211]. In line with the above results, we found that knock out of *MSH2* or *MSH3* mismatch repair (MMR) genes, which are reported to act in the recognition of 3' tails and recruitment of *RAD1-10,* [211, 213], decreased the frequency of His^+ colonies by a factor of 10. This reduction in the frequency of His^+ colonies in wild-type RNase H cells was not seen upon knock out of *MSH6* (**Figure 6.3B**), which functions in MMR but is not involved in 3' tail removal [214]. Markedly different results were obtained when these MMR genes were deleted in the *rnh1 rnh201* background. Knock out of *MSH3* or *MSH6* in a *rnh1 rnh201* background reduced the His^+ frequency by a factor of 2, while deletion of *MSH2* reduced the frequency over a

factor of 9 (**Figure 6.3B**). We could attribute this differential effect to the more prevalent MMR activity of hetroduplex rejection to block homeologous recombination [215] in favor of R-TDR, in combination with the clipping function required for repair by cDNA. In addition, mismatch repair activity may aid in R-TDR if errors occur during DNA synthesis from an RNA template.

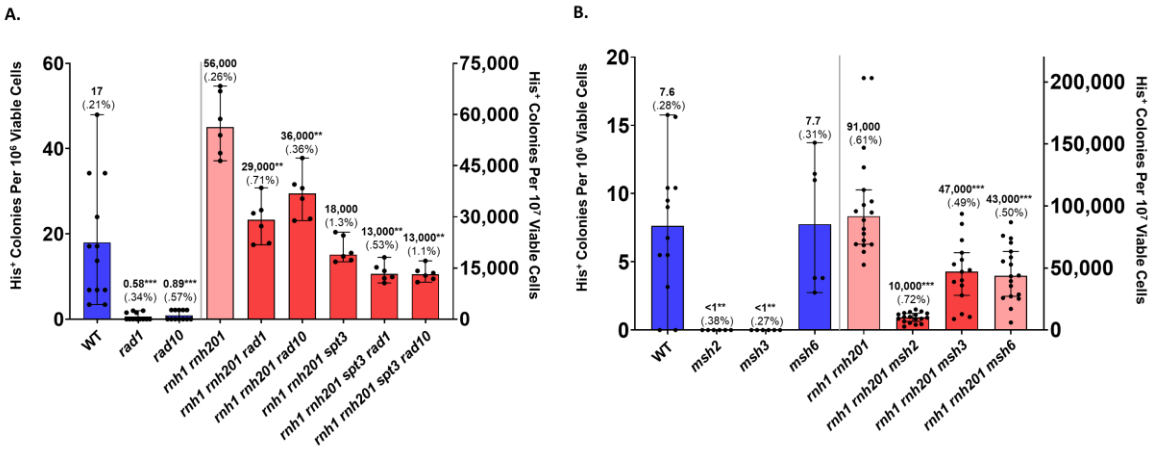


Figure 6.3 Rad1-10 and Msh2-3 are dispensable for R-TDR.

(A) Fluctuation assay showing frequency of His^+ colonies per 10^6 or 10^7 viable cells following DSB induction in clippase mutants. Individual frequencies are plotted. Bars represent mean with range (WT, *rad1*, *rad10) or* median with 95% confidence interval (*rnh1 rnh201, rnh1 rnh201 rad1, rnh1 rnh201 rad10, rnh1 rnh201 spt3, rnh1 rnh201 spt3 rad1, rnh1 rnh201 spt3 rad10*). The genotype of the samples is indicated under each bar. The median or mean of each genotype described above is shown above each bar and survival following DSB is shown in parentheses; N=6-12. *P*-values are shown in (**Table B.6**). (B) Fluctuation assay showing frequency of His^+ colonies per 10^6 or 10^7 viable cells following DSB induction in mismatch repair mutants. Individual frequencies are plotted. Bars represent mean with range (WT, *msh2*, *msh3*, *msh6) or* median with 95% confidence interval (*rnh1 rnh201, rnh1 rnh201 msh2, rnh1 rnh201 msh3, rnh1 rnh201 msh6*). The genotype of the samples is indicated under each bar. The median or mean of each genotype described above is shown above each bar and survival following DSB is shown in parentheses; N=6-18. *P*-values are shown in (**Table B.7**).

6.4.4 R-TDR is independent of the Ty retrotransposon

To determine whether R-TDR requires the presence of the Ty retrotransposon for DNA repair synthesis by Ty reverse transcriptase (RT) on the RNA template, we cloned the *cis* system described in (**Figure 6.1A**), containing an I-SceI endonuclease cut site in place of the HO cut site, onto a yeast centromeric plasmid (HKb-67). We then engineered a *Saccharomyces paradoxus* strain lacking endogenous Ty activity [204] (Ty-less, DG-2204, **Table B.1**) with an integrated copy of the I-SceI endonuclease gene under the *pGAL1* promoter. In this strain (HK-692, 696), we constructed the *rnh1*, *rnh201*, or *rnh1 rnh201* null mutation/s. As a control, the same genetic engineering was done in an *S. cerevisiae* strain containing active Ty (FRO-767), in which we also constructed the knock out mutants *rnh1*, *rnh201*, or *rnh1 rnh201*, as well as *spt3* and *rnh1 rnh201 spt3* (**Table B.1**). We then introduced the HKb-67 plasmid into these *S. paradoxus* and *S. cerevisiae* strains. The presence or absence of the Ty RT protein in the wild-type *S. cerevisiae* and *S. paradoxus* strains (HK-809, 812 and HK-815, 817) (**Table B.1**) was confirmed by western blot analysis (**Figure 6.4A**). *S. cerevisiae* and *S. paradoxus* cells of all the above genotypes containing the *cis* system on plasmid HKb-67 were plated on galactose medium to induce the DSB in *his3*. We then examined the frequency of His^+ colonies formed for these strains. Results obtained for all *S. cerevisiae* strains containing the plasmid-*cis* system with I-SceI endonuclease were in line with published data [40] and data shown in (**Figure 1B**). *S. paradoxus* Ty-less wild-type, *rnh1* or *rnh201* strains behaved similarly to the corresponding *S. cerevisiae* strains containing the *spt3*-null allele (**Figure 6.4B**). We were unable to detect any His^+ colony in these strains, demonstrating that C-TDR does not occur in Ty-less strains. On the contrary, His^+ colonies were detected in the *S. paradoxus* strain containing *rnh1 rnh201* mutations. This is similar to the *S. cerevisiae rnh1 rnh201 spt3*

strain (**Figure 6.4B**). These results demonstrate that while C-TDR requires Ty, the Ty retrotransposon is not required for R-TDR, and R-TDR must proceed with DNA polymerase.

To further validate the role of the Ty retrotransposon in C-TDR and not in R-TDR, we overexpressed a Ty1 element from the *pGAL1* promoter and introduced it in our strains either on a plasmid (pGTyClaI) or by integrating it into the yeast genome. We hypothesized that the overexpression of Ty would increase repair by cDNA but not RNA. We confirmed Ty overexpression by western blot (**Figure 6.4C**). Indeed, overexpression of Ty from pGTyClaI in the *pGAL1* (**Figure B.3C**) or *pTEF* (**Figure B.3D**) system, and from the integrated Ty in the *pTEF* system (**Figure 6.4D**) strongly increased levels of C-TDR**,** but had no impact on DSB repair by DNA oligos (**Figure B.3E**). This suggests specificity of Ty RT to convert RNA into cDNA in our DSB repair assay but not to amplify the sequence of DNA oligos. Moreover, we observed that the overexpression of Ty did not promote R-TDR, because it did not result in increased His^+ frequency in a *rnh1 rnh201 rad1* background that is deficient in repair by cDNA (**Figure 6.4D**). While C-TDR was inhibited over a factor of 25 by loss of *rad1* in wild-type cells when Ty was overexpressed, only less than a factor of two decrease of the His^+ frequency was observed in *rnh1 rnh201 rad1* cells overexpressing Ty (**Figure 6.4D**). Overall, these results demonstrate that C-TDR is Ty RT-driven, while R-TDR does not proceed using a bona fide reverse transcriptase like Ty RT.

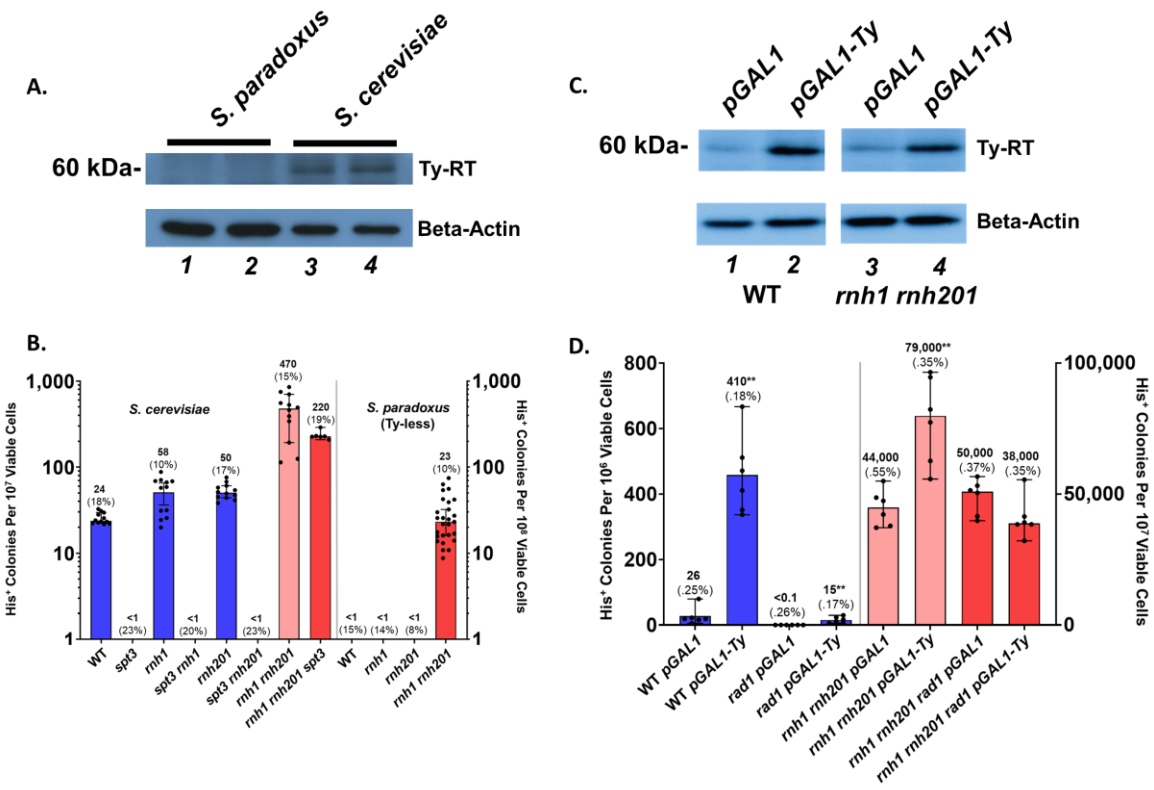


Figure 6.4 C-TDR is driven by the Ty reverse transcriptase.

(A) Western blot showing Ty RT levels using the B8 antibody in exponentially growing *S. paradoxus* and *S. cerevisiae* cells of strains HK-809 and HK-812 in lanes 1 and 2, and strains HK-815 and HK-817 in lanes 3 and 4, respectively. Two independent isolates for each strain are shown. Sample volumes were adjusted to Beta-Actin protein levels. Beta-Actin was used as a housekeeping reference gene. (B) Fluctuation assay of His^+ colonies per 10^7 or 10^8 viable cells following DSB induction comparing frequencies in *S. cerevisiae* and *S. paradoxus* (Ty-less) cells. Individual frequencies are plotted. Bars represent median with 95% confidence interval on a log scale. The genotype of the samples is indicated under each bar. The median of each genotype described above is shown above each bar and survival following DSB is shown in parentheses; N=6-36. *P*-values are shown in **Table B.8**. (C) Western blot showing Ty RT levels using the B8 antibody in strains with *pGAL1* or *pGAL1-Ty*. Sample volumes were adjusted to Beta-Actin protein levels. Beta-Actin was used as a housekeeping reference gene. Strains used are CM-1095 and CM-1099 in lanes 1 and 2 with wild-type RNase H, and CM-1169 and CM-1171 in lanes 3 and 4 with *rnh1 rnh201*-null genotype. (D) Fluctuation assay showing frequency of His^+ colonies per 10^6 or 10^7 viable cells following DSB induction in cells containing integrated *pGAL1-Ty*. Individual frequencies are plotted. Bars represent mean with range (WT *pGAL1*, WT *pGAL1-Ty, rad1 pGAL1, rad1 pGAL1-Ty)* or median with 95% confidence interval (*rnh1 rnh201 pGAL1, rnh1 rnh201 pGAL1-Ty, rnh1 rnh201 rad1 pGAL1, rnh1 rnh201 rad1*

pGAL1-Ty). The genotype of the samples is indicated under each bar. The median or mean of each genotype described above is shown above each bar and survival following DSB is shown in parentheses N=6. *P*-values are shown in **Table B.9**.

6.4.5 DNA polymerase ζ promotes RNA-templated DSB repair

We next sought to determine which DNA polymerase/s is responsible for R-TDR. DNA polymerase δ plays a major role in DSB repair by homologous recombination [216]. Primer extension experiments showed that DNA Pol δ contains some reverse transcriptase activity but has low processivity on RNA templates [45]. In yeast, there are four specialized polymerases that are associated with replication of damaged DNA: Pol4, η, ζ and deoxycytidyl transferase encoded by the *REV1* gene, which forms a complex with Pol ζ subunits [217] [218]. Pol4 works in NHEJ [216], while the translesion polymerases η, ζ and Rev1 can bypass a variety of unnatural or modified nucleotides, including ribonucleotide tracts in DNA [219-221]. *RAD5* has recently been shown to recruit DNA Pol ζ to repair ssDNA gaps at stressed DNA replication forks [222]. Here we investigated the role of these specialized polymerases in R-TDR.

Genetic disruptions showed that elimination of the translesion DNA polymerase ζ pathway (*rev3*, *rev1* or *rad5*-null mutation drops the frequency of R-TDR by a factor of 2.5 in a *rnh1 rnh201 rad1* background, in which cDNA-templated DSB repair (C-TDR) is strongly impaired (**Figure 6.5A**). Loss of Pol η (*rad30-null*) shows no impact on R-TDR in the *rnh1 rnh201 rad1* background (**Figure 6.5A**). Surprisingly, elimination of *pol32,* an accessory subunit of Pol δ and ζ showed only minor decreases in His^+ frequency in the same genetic background. In line with results shown in (**Figure 6.2B**), knock out of the NHEJ-associated Pol Pol4 (*pol4*-null*)* elevated the frequency of R-TDR a factor of 5 (**Figure 6.5A**). These findings support a predominant role of Pol ζ in R-TDR. To examine

whether Pol ζ catalytic activity was responsible for R-TDR, we constructed a low fidelity mutant of the catalytic subunit *rev3* L979F [223]. The low fidelity mutant of Pol ζ showed a minor decrease in the frequency of repair as compared to wild-type Pol ζ (**Figure 6.5B**). To examine whether overexpression of the catalytic subunit of Pol ζ, *REV3*, would increase the frequency of R-TDR, we integrated a copy of the *REV3* gene under the galactose inducible promoter *pGAL1* at the *CAN1* locus in *rnh1 rnh201* cells (**Table B.1**). Indeed, overexpression of *REV3* gene resulted in a modest but significant increase in the frequency of His^+ colonies (**Figure 6.5C**). These data show that overexpression of the catalytic subunit of Pol ζ promotes R-TDR. Overall, our results support a dominant function of Pol ζ in R-TDR.

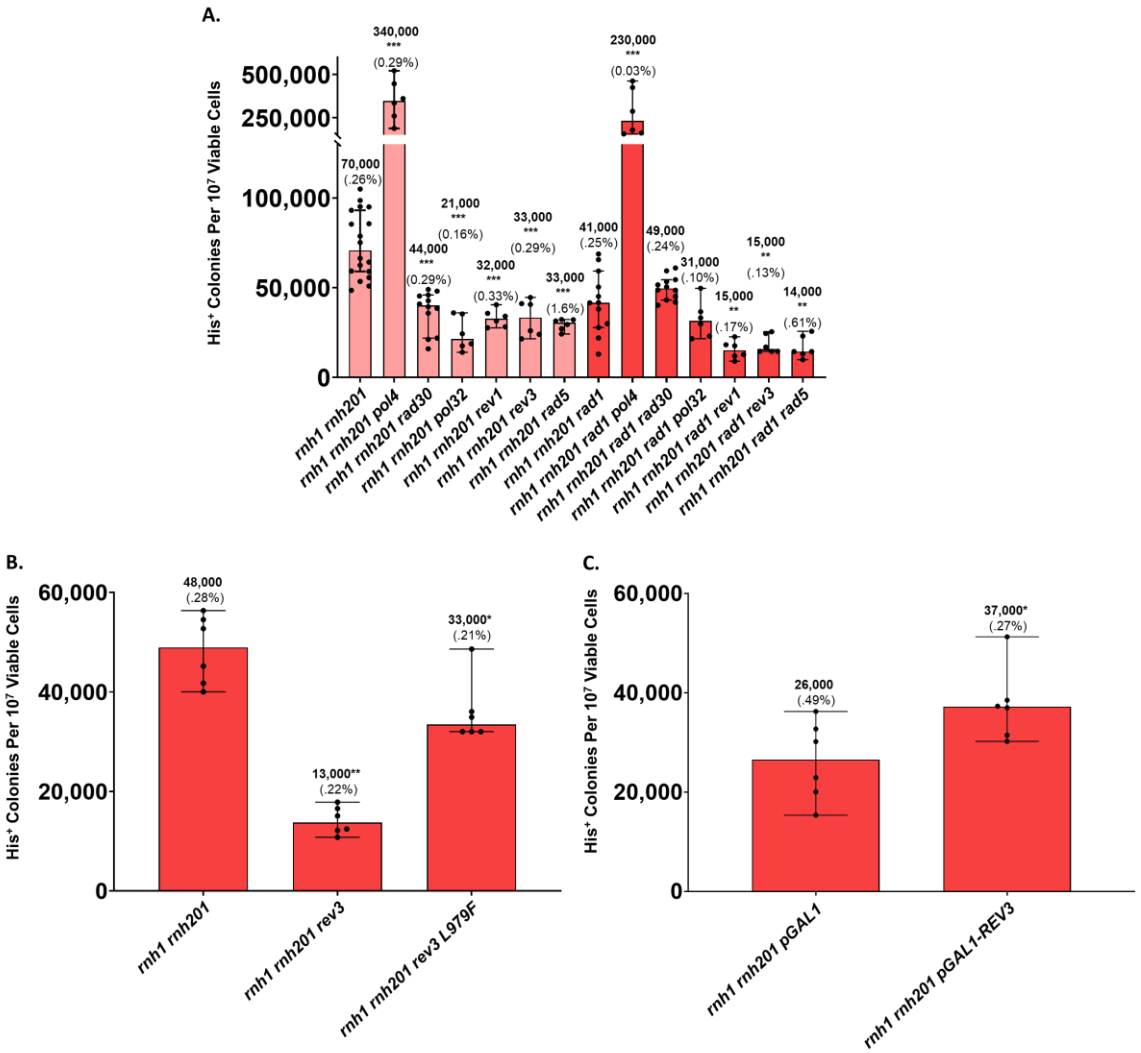


Figure 6.5 DNA polymerase ζ promotes RNA-DNA recombination triggered by a DSB.

(A) Fluctuation assay showing frequency of His$^+$ colonies per 10^7 viable cells following DSB induction in yeast non-essential DNA polymerase mutants. Individual frequencies are plotted. Bars represent median with 95% confidence interval. The genotype of the samples is indicated under each bar. The median is shown above each bar and survival following DSB is shown in parentheses; N=6-18. *P*-values are shown in **Table B.10**. (B) Fluctuation assay showing frequency of His$^+$ colonies per 10^7 viable cells following DSB induction in *rev3* L979F low fidelity mutant. Individual frequencies are plotted. Bars represent median with 95% confidence interval. The genotype of the samples is indicated under each bar. The median is shown above each bar and survival following DSB is shown in parentheses; N=6. *P*-values *rnh1 rnh201 - rnh1 rnh201 rev3* (0.0022), *rnh1 rnh201 - rnh1 rnh201 rev3 L979F* (0.0152). (C) Fluctuation assay showing frequency of His$^+$ colonies per 10^7 viable cells following DSB induction in cells overexpressing *REV3*. Individual frequencies are plotted. Bars represent median with 95% confidence interval. The genotype of the samples

is indicated under each bar. The median is shown above each bar and survival following DSB is shown in parentheses; N=6. *P*-values *rnh1 rnh201 pGAL1 - rnh1 rnh201 pGAL1-REV3* (0.0260).

6.4.6 Rad52-independent RNA-DNA recombination in the absence of an induced DSB

Upon replacement of the inducible *pGAL1* with the constitutive *pTEF* promoter, we unexpectedly detected abundant formation of His^+ colonies in the *rnh1 rnh201*, *rnh1 rnh201 spt3* and *rnh1 rnh201 rad1* backgrounds without induction of the DSB (**Figure 6.6A**). His^+ colonies were detected by growing cells in glucose containing medium, suppressing activation of the HO endonuclease driven by the *pGAL1* promoter. This observation was not the result of leaky HO endonuclease expression in glucose medium as knock out of the *HO* endonuclease gene did not reduce the frequency of His^+ colonies in similar conditions (**Figure 6.6B**). We hypothesized that formation of an RNA-DNA hybrid between the antisense *his3* transcript and the *his3* DNA gene may recruit mismatch or nucleotide excision repair nucleases resulting in cleavage within the *his3* locus, and recombination by RNA. We investigated the role of previously reported nucleases (*MLH1* and *RAD1*) involved in processing of R-loops [224-226] and found no difference in the frequency of His^+ colonies (**Figure 6.6C**). We cannot eliminate a redundant role of other nucleases or spontaneous cleavage, but these results demonstrate the capacity of RNA to modify DNA even in the absence of an induced DSB at a homologous locus. We termed this mechanism as RNA-templated DNA modification (R-TDM).

Similar to R-TDR, R-TDM is independent of cDNA-mediated repair because loss of *SPT3* or *RAD1* does not decrease the His^+ frequency in the *cis* assay in the absence of the induced DSB in *his3* (**Figures 6.6A and B**). Additionally, overexpression of Ty in strains lacking the *HO* gene (*ho*-null) did not result in increased frequency of His^+ colonies

either in the wild-type or *rnh1 rnh201* backgrounds (**Figure 6.6D**). We hypothesized that the R-TDM mechanism, which is independent of a DSB, requires formation of an RNA-DNA hybrid (R-loops) during transcription [227]. Because the Rad52 protein is strongly required for DSB repair by RNA in R-TDR by promoting an inverse strand exchange reaction between the broken DNA ends and the single-stranded RNA template [47], we examined whether Rad52 was required for R-TDM. We tested the ability of Rad52 to stimulate R-loop formation *in vitro* and found that while Rad52 can promote D-loop formation with DNA, it cannot promote R-loop formation with RNA (**Figure 6.7**). In line with the biochemical data, loss of genes associated with homologous recombination (*rad52* or *rad59*-null) were inconsequential to the frequency of R-TDM in the yeast cells (**Figure 6.6B**). Only *rad51*– null shows a factor of 2 increase in the frequency of R-TDM. These results demonstrate that in the absence of an induced DSB, transcript RNA has the capacity to recombine with homologous DNA sequences and mediate DNA modifications even without the catalytic support of a recombination protein. The data also uncover a novel mechanism of Rad52-independent recombination in yeast.

6.4.7 DNA polymerase ζ is essential for RNA-DNA recombination in the absence of an induced DSB

While Ty RT is not required for R-TDM, as shown by knockout of *SPT3* or overexpression of Ty (**Figure 6.6 B and D**), Pol ζ is essential. Knock out of the *REV3* gene strongly reduces the frequency of His^+ colonies in *rnh1 rnh201 spt3* cells grown in glucose (**Figure 6.6A**). In line with R-TDR results, loss of the DNA Pol ζ translesion synthesis pathways (*rev1*, *rev3* or *rad5*-null) strongly reduced the frequency of R-TDM (**Figure 6.6E**). Furthermore, the low fidelity DNA synthesis derivative *rev3*-L979F strongly

diminishes the frequency of His^+ colonies in *rnh1 rnh201* cells grown in glucose with no DSB induction (**Figure 6.6E**). Sequencing of the *HIS3* locus from several His^+ colonies isolated from wild-type *REV3* found 30/30 with correct *HIS3* sequence. In contrast, only 15/18 in the *rev3*-L979F mutant had correct *HIS3* sequence (**Figure B.4**). This result supports synthesis by Pol ζ in R-TDM. Lastly, overexpression of the *REV3* gene coding for the catalytic subunit of Pol ζ results in an increase (a factor of two) in the frequency of His^+ colonies in *rnh1 rnh201* cells lacking the *HO* gene (*ho*-null) (**Figure 6.6F**). Overall, these findings demonstrate an essential function of Pol ζ in R-TDM.

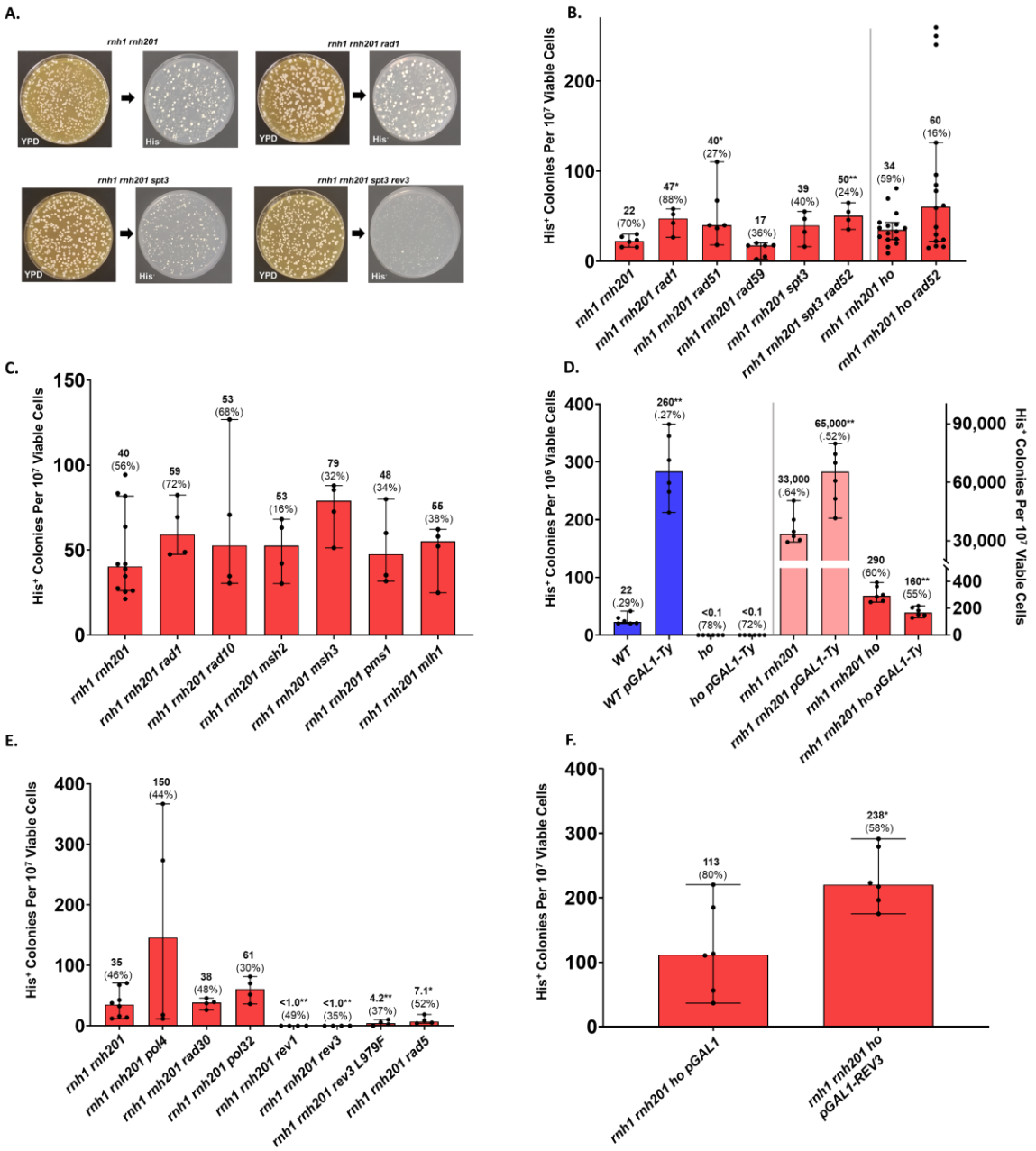


Figure 6.6 R-TDM requires DNA polymerase ζ.

(A) Representative plates showing single colony isolates of *rnh1 rnh201*, *rnh1 rnh201 rad1*, *rnh1 rnh201 spt3* and *rnh1 rnh201 spt3 rev3* mutant strains on YPD medium (no DSB induction) that were replica-plated onto His$^-$ medium. (B) Fluctuation assay showing frequency of His$^+$ colonies per 10^7 viable cells with no DSB induction for cells of different recombination mutants. Individual frequencies are plotted. Bars represent median with 95% confidence interval. The genotype of the samples is indicated under each bar. The median is shown above each bar and survival shown in parentheses; N=4-16. *P*-values are shown in **Table B.11**. (C) Fluctuation assay showing frequency of His$^+$ colonies per 10^7

viable cells with no DSB induction for cells of clippase and mismatch excision repair mutants. Individual frequencies are plotted. Bars represent median with 95% confidence interval. The genotype of the samples is indicated under each bar. The median is shown above each bar and survival shown in parentheses; N=4-12. *P*-values are shown in **Table B.12**. (D) Fluctuation assay showing frequency of His^+ colonies per 10^6 or 10^7 viable cells with and without DSB induction in strains overexpressing the Ty transposon. Individual frequencies are plotted. Bars represent median with 95% confidence interval. The genotype of the samples is indicated under each bar. The median is shown above each bar and survival shown in parentheses; N=6 *P*-values are shown in **Table B.13**. (E) Fluctuation assay showing frequency of His^+ colonies per 10^7 viable cells with no DSB induction for cells of non-essential DNA polymerase mutants. Individual frequencies are plotted. Bars represent median with 95% confidence interval. The genotype of the samples is indicated under each bar. The median is shown above each bar and survival shown in parentheses; N=4-8 *P*-values are shown in **Table B.14**. (F) Fluctuation assay showing frequency of His^+ colonies per 10^7 viable cells with no DSB induction for cells overexpressing *REV3*. Individual frequencies are plotted. Bars represent median with 95% confidence interval. The genotype of the samples is indicated under each bar. The median is shown above each bar and survival shown in parentheses; N=6 *P*-value *rnh1 rnh201 ho pGAL1 - rnh1 rnh201 ho pGAL-REV3* (0.0260).

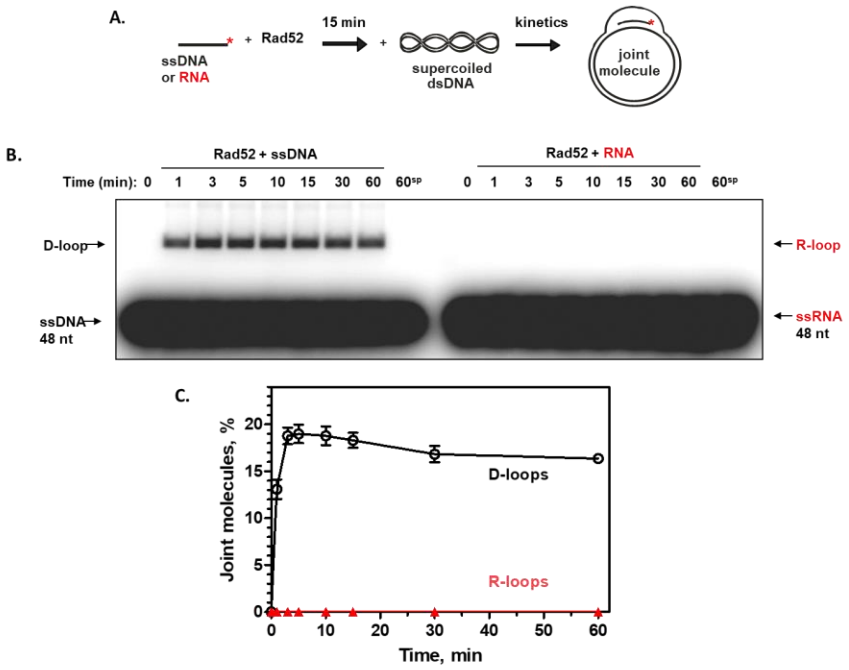


Figure 6.7 ScRad52 promotes D-loop but not R-loop formation.

(A) The reaction schemes. A red asterisk denotes the ^{32}P label. (B) The kinetics of D- and R-loop formation. An example of three repeats is shown. ScRad52 protein (450 nM) was preincubated with a 48-mer ^{32}P- labeled ssDNA (no. 211; 3 µM, nt) or with a 48-mer RNA of identical sequence (no. 501; 3 µM, nt) for 15 min at 37 °C, and D-loop formation was initiated by addition of homologous pUC19 dsDNA (67.2 µM, nt). The reaction products at indicated time points were analyzed by electrophoresis in a 1% agarose gel. In the

reaction marked by "60^{sp}", ScRad52 was replaced with storage buffer, and the reaction was carried out for 60 min at 37°C. (C) Data from B represented as a graph. Error bars indicate the standard error of the mean (SEM). N=3

6.5 Discussion

6.5.1 Mechanism of RNA-templated DSB repair (R-TDR) driven by Pol ζ

Understanding the potential of RNA to recombine with DNA has been difficult given RNA sequence resemblance to the DNA sequence from which it is generated. However, large scale RNA-mediated genome rearrangements have been observed in ciliates [42, 44]. In addition, it has been proposed that RNA-templated mechanisms may drive somatic hypermutation [41]. We previously demonstrated that RNA is recombinogenic in the absence of RNase H and is aided by the Rad52 function of annealing RNA to DNA by an inverse strand exchange reaction [40, 47]. In addition, this process does not require extensive end resection [47].

Here we expanded on this model to find limited dependence on, *RAD1-10*, *MSH2-3* clippase suggesting R-TDR can proceed with limited 3'-nonhomologous tails removal. We hypothesize that extensive resection is favored in C-TDR, facilitating recombination between the broken DNA ends and the cDNA molecule. In contrast, during R-TDR, extensive resection is not needed as the donor RNA is already localized to the DSB site and can be efficiently annealed to RNA even at blunt ended DNA structures by Rad52 in an inverse strand exchange reaction [47]. Removal of the 3' flap generated in the artificial intron could be attributed to limited tail remove by the 3'-to-5' proofreading activity of DNA polymerase delta [214]. We exploit this difference in the preference for more

extensive resection for C-TDR vs limited resection in R-TDR to distinguish between these repair pathways. This is in conjunction with *spt3*-null mutants, which inhibit transcription of endogenous Ty elements (**Figure 6.8A**). Recently, Rad52 was shown to also limit resection in budding yeast [228], which may aid in R-TDR. Surprisingly, we found little effect of *rad59*-null or *rad51*-null mutants in R-TDR when the antisense RNA is driven by the *pTEF* promoter. This is different from our previous work exploiting the galactose inducible donor system showing a stimulation in the frequency of R-TDR in the absence of either *rad51* or *rad59* [40, 47]. We propose that loss of either *rad51* or *rad59* inhibits competition with sister chromatid recombination. However, this competition may be less favorable for the sister chromatid in the *pTEF* constitutive system, as the donor RNA is already actively transcribed at the time the DSB is induced. In contrast to effects seen in recombination mutants, NHEJ mutants show a dramatically increased frequency of R-TDR events but a marked loss of cell viability following DSB induction.

Remarkably, we discovered a unique role for the translesion DNA polymerase ζ in mediating R-TDR. The current understanding of DNA translesion synthesis pathways in yeast suggests that replication stalling results in ubiquitination of the PCNA clamp loader, switching replicative polymerases to translesion polymerases (Pol η, ζ) and bypass of the damaged DNA [229]. We find reductions in the frequency of R-TDR in the absence of translesion polymerases, predominantly DNA Pol ζ in a *rnh1 rnh201 rad1* background, in which C-TDR is inactive. Recombination-associated DNA synthesis initially proceeds with high fidelity replicative polymerase δ [230-232] but translesion DNA polymerases are also associated with DSB repair [233-235]. Earlier work has shown that yeast replicative polymerases contain minimal reverse transcriptase activity [45], in addition to documented

reverse transcriptase activity of human Pol η [220] and yeast Pol ζ in the bypass of multiple embedded ribonucleotides in DNA [219]. However, the fidelity of DNA polymerase ζ is known to be significantly lower than that of replicative polymerases with a preference for base substitutions [236]. Interestingly, this supports our results showing *MSH6* is important for R-TDR, possibly in the repair of mismatches generated by DNA Polymerase ζ but not C-TDR, which is Pol ζ-independent. We propose that during R-TDR, DNA polymerase δ encounters donor RNA annealed to the 3' end of the DSB, this results in polymerase switching to DNA Pol ζ and provides increased reverse transcriptase capabilities driving repair of the DSB aided directly by an RNA template **(Figure 6. 8A)**.

6.5.2 Mechanism of RNA-mediated DNA modification (R-TDM)

In addition to the ability of RNA to transfer information back to DNA following a DSB, we found that RNA can transfer genetic information back to DNA in the absence of RNase H without the induction of a DSB. This process is independent of the major recombination proteins in yeast, Rad52, and Rad59, and stimulated in the absence of Rad51 highlighting a unique form of recombination in yeast. We hypothesize that in the absence of RNase H, R-loops form with the transcript RNA that can recombine with DNA in a Rad52-independent manner **(Figure 6.8B)**. This RNA-DNA recombination in the absence of an induced DSB is strongly dependent on DNA Pol ζ translesion synthesis pathway (*REV1*, *REV3* and *RAD5*). We detected no role of nucleases involved in cleavage of R-loops structures (*RAD1* or *MLH1*) [226]. However, we cannot eliminate the role of other nuclease, spontaneous nicks or DSBs. In addition, reports have suggested a possible role of R-loops in the initiation of origin-independent replication events in *E. coli* [237], yeast [238] and human mitochondria [239]. Possibly, the donor RNA or fragments of the donor

RNA may be incorporated during DNA synthesis leading to RNA-mediated DNA modification events similar to oligonucleotide incorporation during replication [240].

6.5.3 Mechanism of cDNA-templated DSB repair (C-TDR)

Mobile genetic elements and their impact on genome diversification are areas of intense studies [241]. However, information on how these mobile elements affect genome stability is scarce. Reports have demonstrated cDNA-mediated DSB repair events in yeast, mice and human cells [40, 48, 57, 242, 243]. Others have shown a role of cDNA recombination events as drivers of copy number variations in human neuronal cells, leading to a speculative mechanism of "recording" and "playback" of preferred gene variants [244]. Previous reports have demonstrated strong requirements for *RAD52* in cDNA-mediated recombination between Ty elements in yeast, but found that *RAD1* was not important for this process [245]. This is contrary to our results showing a strong dependence on *RAD1-10* for cDNA-templated recombination between genomic *his3* and *HIS3* cDNA. We also find a strong dependence on *MSH2-3* for cDNA-templated DSB repair but not *MSH6*. This likely highlights a requirement for removal of nonhomologous ends during recombination. Removal of intronic sequence contained in the DNA is required for the formation of His^+ colonies and this likely requires the activity of *RAD1-10* and *MSH2-3* (**Figure 6.8C**). We show that by stimulating the production of Ty1 *in vivo*, the frequency of cDNA-templated DSB repair increases, suggesting an increase in *HIS3* cDNA, but it is unknown how this *his3* RNA is captured and reverse transcribed by Ty. Studies of Ty reverse transcriptase have indicated Ty1 reverse transcriptase transferring from normal Ty1 template ends to various tRNA templates [246]. This has also been reported for non-LTR retroelements like those of LINE-1 elements in humans to poly-A tracts [39, 247]. It would be interesting to

understand the mechanism and rules which govern cDNA amplification of genomic transcripts, and their role on genome in/stability. We show that RNA molecules can modify genomic the DNA sequence at a homologous genomic locus in yeast DNA through a variety of different mechanisms interacting with multiple DNA repair pathways. This illustrates a powerful role of RNA in directly templating genomic modifications as a driver of genome in/stability.

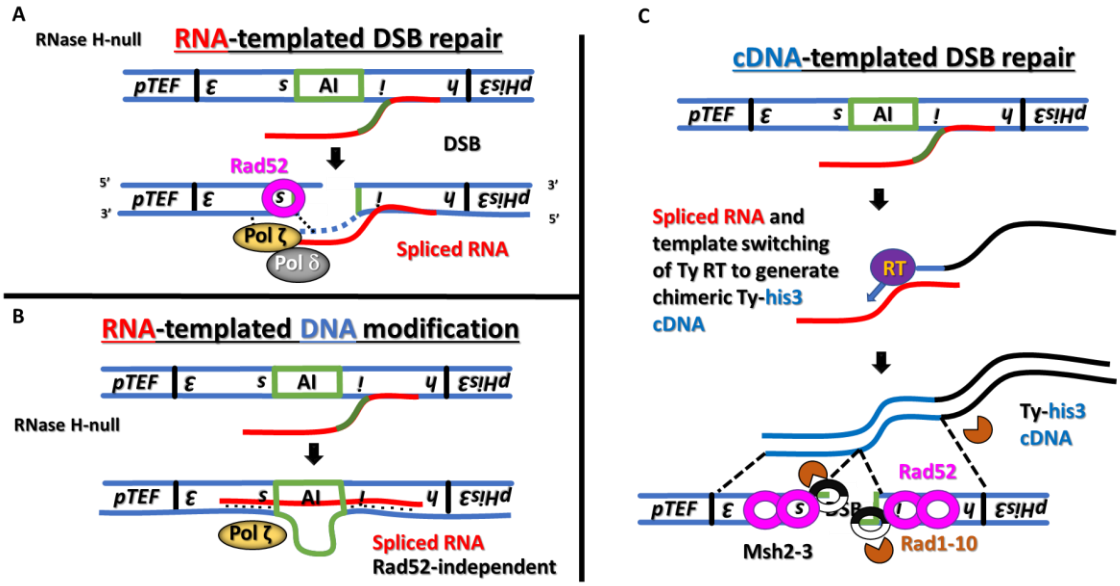


Figure 6.8 Model of RNA-templated DSB repair (R-TDR), RNA-templated DNA modification (R-TDM) and cDNA-templated DSB repair (C-TDR).

(A) Transcript-RNA anneals back to DNA following DSB aided by Rad52. RNA is then used as a template for DSB repair synthesis, aided by DNA Pol ζ. (B) Transcript-RNA forms an R-loop structure without the need of Rad52. RNA is then used as a template by DNA Pol ζ to guide removal of the intronic sequence in DNA. (C) Spliced antisense *his3* RNA is reverse transcribed by Ty reverse transcriptase and used as a templated to facilitate repair of DSB inside of the *his3* locus from which the RNA was generated. This process requires Rad52 for annealing, and *MSH2-3* along with *RAD1-10* in the removal of non-homologous tail DNA.

6.6 Acknowledgements

We thank D. Garfinkel for plasmids pBDG102, pGTyClaI and for strain *Saccharomyces paradoxus* DG-2204. K.D. Koh for strain construction of KK strains and Y. Shen for YS strains. We thank E. Graf and D. Sas for their technical assistance. We thank H. Ghalei, S. Marsili and K. Mukherjee for critical reading of the manuscript, and all the people of the Storici lab for assistance and feedback on this research. We acknowledge funding from the National Cancer Institute of the NIH (grant numbers CA188347 and P30CA056036 to A.V.M.), Drexel Coulter Program Award (to A.V.M.), the National Institute of General Medical Sciences (NIGMS) of the NIH (grant number GM115927 to F.S.), the National Science Foundation fund (grant number 1615335 to F.S.), and the Howard Hughes Medical Institute Faculty Scholar (grant number 55108574 to F.S.) for supporting this work.

6.7 Declaration of Interests

The authors declare no competing financial interest.

CHAPTER 7. CONCLUSION

Biochemical assays have suggested between 70-90% of the human genome is transcribed including both mRNA and ncRNA [30]. This implies that large parts of genetic information are transferred from DNA to RNA, but can this information be transferred back from RNA to DNA? We show that RNA molecules can transfer information back to DNA by a variety of different mechanisms that are mediated by different molecular pathways. These processes highlight the cross talk between nucleic acids and provide possible insights into genome evolution and stability. We highlight the molecular pathways in budding yeast by which RNA can drive genomic changes and provide future directions for RNA-mediated DNA modifications.

7.1 Transposon-mediated double-strand break repair.

Mobile genetic elements are widely recognized as important drivers of genetic diversity [248]. The process of transposition is largely inhibited by endogenous cellular pathways to prevent excessive genome rearrangements [249]. However, DNA damage can stimulate genome instability by transposition events [250]. The Ty transposon of yeast have been shown to transpose into sites of DSB [55], as have L1 elements in human [38]. These transposons can also capture endogenous transcript RNA and reverse transcribe them into cDNA molecules that can be captured at sites of DSBs [242, 243]. We show that endogenous transcript RNA captured by Ty transposons can be used as a template for DSB repair in a homologous recombination mechanism (**Figure 6.8C**). This process requires homologous recombination factors, along with DNA clippases. A cDNA-mediated DNA recombination mechanism is postulated in human neurons which show thousands of variant

genomic cDNAs recombination events identified by their lack of intronic sequence [244]. These neuronal recombination events are hypothesized to provide a "recoding" of neural activity for selective "playback" of preferred gene variants [244].

- **Ty retrotransposon stimulates cDNA-mediated DNA repair.**
- **Transposon-mediated cDNA DSB repair is repaired by a homologous recombination mechanism**
- **Transposon-mediated DSB repair requires DNA clippase *RAD1-10* and *MSH2-3***

7.2 Transcript RNA-templated DSB repair

Transcript RNA is homologous to the DNA from which it was generated from. This has made it difficult to determine RNAs ability to modify genomic DNA. It has been proposed that RNA-DNA recombination with error prone reverse transcription or RNA editing events followed by reverse transcription may be responsible for immunoglobulin somatic hypermutation [41, 251]. Furthermore, an RNA-templated genome rearrangement process seems to have been adopted in the development of the ciliate *Oxytricha trifallax* genome, as long RNA transcripts have been shown to mediate large scale genome rearrangements [44] and this process requires the activity of endogenous transposons [252]. We show that transcript RNA can be used in a more general form of DSB repair. We find that in the absence of RNase H function, transcript RNA can interact with Rad52 to facilitate annealing of RNA to DNA and drive a homologous recombination mechanism with RNA. We hypothesize that during RNA-templated DSB repair, following annealing of RNA to DNA by Rad52, stalling of Pol δ on an RNA template leads to polymerase

switching to Pol ζ, which can provide limited reverse transcriptase activity and repair of DSB by RNA. Contrary to repair by cDNA, which requires DNA clippase Rad1-10 and Msh2-3, we find limited dependence in RNA-templated DSB repair. Interestingly, *MSH6*, involved in mismatch repair shows no impact in repair by cDNA but minor impacts in RNA-templated repair, possibly through the incorporation of errors by Pol ζ. We have previously shown that RNA-templated DSB repair prefers limited DNA end resection at the site of the DSB [47]. This supports our findings of a histone H1 linker *HHO1* independent process, which instead functions in recombination at regions with extensive non homologous tails. (**Figure 6.8A**).

- **Transcript-RNA templated DSB repair is independent of Ty retrotransposon**
- **Transcript-RNA templated DSB repair is aided by translesion polymerase ζ**
- **Transcript-RNA templated DSB repair occurs in a homologous recombination manner and is independent of NHEJ**

7.3 Transcript RNA-templated DNA modification

We surprisingly find that donor RNA can template the removal of intronic sequences in the absence of an induced DSB. We termed this process as RNA-templated DNA modification. RNA-templated DNA modification does not require the recombination proteins Rad52, Rad51 or Rad59, which are essential for recombination-based mechanisms of DSB repair. Furthermore, RNA-templated DNA modification events are independent of nucleases involved in R-loop processing (*RAD1* & *MLH1*) [226]. In addition, RNA-

templated DNA modification is completely dependent on DNA polymerase ζ, this is different from RNA-templated repair of a DSB, which can proceed without DNA polymerase ζ but with at a much lower frequency, suggesting there may be some redundancy for limited reverse transcriptase activity of DNA polymerases. RNA-templated DNA modification may require more extensive reverse transcriptase activity, which could explain the dependence of DNA polymerase ζ. It is unclear how RNA can modify DNA the absence of a DSB. We find that this process is independent of Ty transposon, suggesting this is not a cDNA mediated mechanism. The independence of recombination proteins and the complete dependence on DNA polymerase ζ vs. a strong reliance on Rad52 and a limited dependence on DNA polymerase ζ with a DSB suggest that repair in the absence of an induced DSB proceeds by a separate mechanism (**Figure 6.8B**). Reports have indicated a role of R-loops in the initiation of origin-independent replication events in *E. coli* [237], yeast [238] and human mitochondria [239]. It is possible that the donor RNA or fragments of the donor RNA may be incorporated during DNA synthesis leading to RNA-mediated DNA modification events similar to oligonucleotide incorporation during replication [240].

- **RNA can directly template genomic modifications in the absence of an induced DSB**
- **RNA-templated DNA modification is strongly dependent on the DNA polymerase ζ pathway**
- **RNA-templated DNA modification is independent of recombination factors *RAD52, RAD51* and *RAD59***

7.4 Future Directions

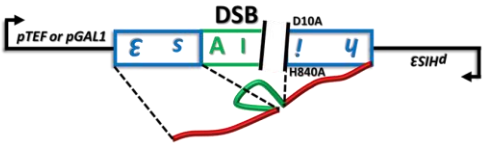
Our work has shown that RNA can modify genomic DNA by a variety of mechanisms. However, there are many questions left to answer. It has been proposed that mobile genetic element domestication played a crucial role in the evolution of different molecular pathways including the CRISPR bacterial immune system [253], telomerase [254] and V(D)J recombination [255]. We show that retrotransposons can amplify generic transcript RNA. What are the rules governing amplification of generic transcripts by endogenous retroelements? Retrotransposons belong to class 1 transposons, which are further divided into two subclasses, LTR (long-terminal repeat) and non-LTR. Yeast Ty transposon belongs to the LTR family of transposons and amplifies it's on RNA by extension of a methionine tRNA following binding at a primer binding site (PBS) inside of the Ty RNA [256]. Can other RNA molecules containing PBS be amplified? Ty reverse transcriptase has been reported to transfer from Ty RNA to various RNA templates [246]. Similar approaches using retron reverse transcriptase to generate ssDNA *in vivo* have been used to convert transient cellular signals into genomically encoded memories in living cells [257]. In addition, non-LTR containing transposon like LINE-1 (L1) elements of humans are abundant, consisting of almost 20% of the entire human genome [258]. Can these L1 elements function to facilitate RNA-templated DSB repair? LINE-1 (L1) elements have been shown to insert at sites of DSBs and contain fragments of endogenous mRNA, suggesting they can accommodate non-L1 RNA templates at DSB sites [38]. Similar results were seen in telomeric regions [38, 61]. In fact, *Drosophila melanogaster* lack a dedicated telomerase enzyme that is compensated by L1 reverse transcriptases [259].

In our current genetic assay to detect RNA-templated DSB repair and modification, we find that RNA-templated events require the removal of RNase H enzymes. This system

depends on a HO endonuclease generated DSB in the middle of an artificial intronic sequence that leaves a 135 bp 3' tail that needs to be removed to facilitate RNA-templated DSB repair. We hypothesized that by generating a DSB at the intron-exon junction, this may allow for detection in the presence of a RNase H. Using the CRISPR/Cas9 system to generate DSBs at the intron-exon junction (**Figure 7.1A**), we do detect repair in the absence of Ty functions by deletion of *spt3* (**Figure 7.1B**). This suggest that RNA-templated DSB repair can function in the presence of RNase H enzyme. Furthermore, we generated Cas9 nickase mutants (**Figures 7.1A, C and D**) in the absence of *rnh1 rnh201* and found that nicking of the DNA stimulates RNA-DNA recombination in comparison to a DSB. We find that nicking of the transcript template strand (H840A) produces a minor increase by a factor of 1.35 relative to a DSB. Surprisingly, nicking of the opposite strand (D10A) stimulated repair by a factor of 3. Could nicks stimulate RNA-DNA recombination in wild-type cells?

Our insights into the mechanisms of RNA-DNA recombination events are the result of a locus specific genetic assay that depends on splicing of an antisense RNA to template repair of a DSB (**Figure 7A**). However, this genetic assay does not detect events of RNA-DNA recombination templated by pre-mRNA, as repair would result in the locus retaining the intronic sequence. Is repair by pre-mRNA more efficient than by spliced RNA? Can single nucleotide polymorphisms (SNPs) be transferred in RNA-DNA recombination? How common are RNA-DNA recombination events genome wide? All these questions require new and novel genetic assays to uncover the global aspects and molecular mechanisms of RNA-driven genetic changes.

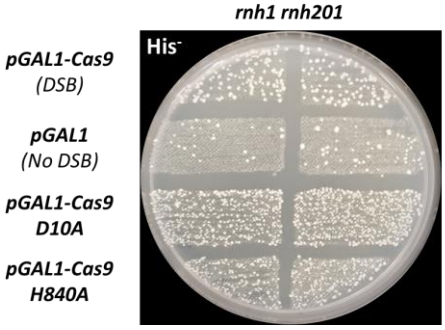
A.



B.

Genotype	His+ frequency per 10^7 viable cells	Survival %
WT	**80** (57-123)	4.4
spt3	**3.5** (<1-7)**	4.4

C.



D.

Genotype- *rnh1 rnh201*	His+ frequency per 10^7 viable cells	Survival %
DSB	**2,000** (1700-3,000)	6
No DSB	**500** (350-800)**	30
D10A	**6,000** (5,000-7,000)**	45
H840A	**2,700** (2,500-3,100)*	50

Figure 7.1 RNA-templated DSB repair by CRISPR/Cas9 induced DSB.

A. Scheme of RNA-templated DSB genetic assay with Cas9 induced DSB or nick at intron-exon junction. Location of nick mutants are listed. B. Frequency of RNA-templated DSB repair by Cas9 DSB in galactose inducible donor system. *P*-value **- 0.0022. C. Example of replica plating results of RNA-templated nick repair in *rnh1 rnh201*-null strains with *pTEF* donor system. D. Frequency of RNA-templated nick repair in *rnh1 rnh201*-null strains with *pTEF* donor system. *P*-value **= 0.0022, *= 0.041.

APPENDIX A. SUPPLEMENTARY MATERIAL FOR CHAPTER 5

Table A.1 List of strains used in this study.

Strain	Genotype	References
BY4741	*MAT*a, *his3Δ1, leu2Δ0, lys2Δ0, ura3Δ0*	
YMT 1001	*BY4741, hho1 T10:: CORE-I-*Sce*I-KANMX4-KIURA3 (GSKU)*	This study
YMT1013	*YMT1001, hho1 T10A*	This study
YMT1015	*YMT1001, hho1 T10E*	This study
YMT1055	*BY4741, trp5::GSKU*	This study
YMT1056	*YMT1013, trp5::GSKU*	This study
YMT1057	*YMT1015, trp5::GSKU*	This study
YMT1067	*BY4741, hho1 S141::GSKU*	This study
YMT1071	*YMT1067, hho1 S141A*	This study
YMT1072	*YMT1067, hho1 S141E*	This study
YMT1075	*YMT1055, hho1D::LEU2*	This study
YMT1076	*BY4741, HHO1-HA*	This study
YMT1083	*BY4741, hho1 S174::GSKU*	This study
YMT1077	*YMT1083, hho1 S174A*	This study
YMT1078	*YMT1083, hho1 S174E*	This study
YMT1079	*YMT1077, trp5::GSKU*	This study
YMT1080	*YMT1078, trp5::GSKU*	This study
YMT1081	*YMT1071, trp5::GSKU*	This study
YMT1082	*YMT1072, trp5::GSKU*	This study
YMT1083	*BY4741, hho1 S173::GSKU*	This study
YMT1090	*YMT1083, S173A*	This study
YMT1091	*YMT1083, S173E*	This study
YMT1096	*BY4741, Hho1 S65::GSKU*	This study
YMT1097	*YMT1096, S65A*	This study
YMT1098	*YMT1096, S65E*	This study
YMT1099	*YMT1097, trp5::GSKU*	This study
YMT1100	*YMT1098, trp5::GSKU*	This study
YMT1101	*YMT1090, trp5::GSKU*	This study
YMT1102	*YMT1091, trp5::GSKU*	This study
YS528	*hoΔ hmlΔ::ADE1 matαΔ::hisG hmrΔ::ADE1 ade1 (leu2::pGAL1mhis3AI-ADE3)::HO lys5 trp1::hisG ura3-52 ade3::GAL::HO (his3::HOcs)::Trp1 YCLWTy2-1Δ rnh1Δ::NAT rnh201Δ::hygMX4 pGAL1Δ::klURA3*	Keskin *et al* Nature 2012
CM280	*YS528, (pGAL1Δ::klURA3)::pTEF*	
CM1012	*CM280, hho1::CORE*	This study
CM1042	*CM1012, hho1 S173A*	This study
CM1043	*CM1012, hho1 S173E*	This study

CM1044	*CM1012, hho1 S174A*	This study
CM1045	*CM1012, hho1 S174E*	This study
YMT1111	*CM280, His3::GSKU*	This study
YMT1112	*CM1064, His3::GSKU*	This study
YMT1115	*CM1043, His3::GSKU*	This study
YMT1116	*CM1044, His3::GSKU*	This study
YMT1117	*CM1045, His3::GSKU*	This study
YMT1118	*MATa/α, hho1 T10A*	This study (M1)
YMT1119	*MATa/α, hho1 T10E*	This study (M1)
YMT1120	*MATa/α, hho1 S65A*	This study (M1)
YMT1121	*MATa/α, hho1 S65E*	This study (M1)
YMT1144	*MATa/α, hho1 S141A*	This study (M1)
YMT1145	*MATa/α, hho1 S141E*	This study (M1)
YMT1122	*MATa/α, hho1 S173A*	This study (M1)
YMT1123	*MATa/α, hho1 S173E*	This study (M1)
YMT1124	*MATa/α, hho1 S174A*	This study (M1)
YMT1125	*MATa/α, hho1 S174E*	This study (M1)
YMT1126	*MATa/α, WT*	This study (M1)
YMT1127	*CM280, his3::CORE*	This study
YMT1128	*CM1044, his3::CORE*	This study
YMT1129	*CM1045, his3::CORE*	This study
YMT1130	*CM280, his3::URA3*	This study
YMT1131	*CM1044, his3::URA3*	This study
YMT1132	*CM1045, his3::URA3*	This study
YMT1146	*MATa/α, hho1 T10A*	This study (M2)
YMT1147	*MATa/α, hho1 T10E*	This study (M2)
YMT1148	*MATa/α, hho1 S65A*	This study (M2)
YMT1149	*MATa/α, hho1 S65E*	This study (M2)
YMT1150	*MATa/α, hho1 S141A*	This study (M2)
YMT1151	*MATa/α, hho1 S141E*	This study (M2)
YMT1152	*MATa/α, hho1 S173A*	This study (M2)
YMT1153	*MATa/α, hho1 S173E*	This study (M2)
YMT1154	*MATa/α, hho1 S174A*	This study (M2)
YMT1155	*MATa/α, hho1 S174E*	This study (M2)
YMT1156	*MATa/α, WT*	This study (M2)

Table A.2 List of oligonucleotides used in this study.

Name	Sequence	Description
KM1	CAACTAAAGCAACATGGCACCCAAGAAATCCACTACCAAGTTCGTACGCTGCAGGTCGAC	Amplify GSKU @ HHO1 T10 F
KM2	CCTTGCCTTTGCTGGTTGCAGGCTTCTTGCCCTTACTTGTCCGCGCGTTGGCCGATTCATT	Amplify GSKU @ HHO1 T10 R
KM3	CAACTAAAGCAACATGGCACCCAAGAAATCCACTACCAAGGCTACAAGTAAGGGCAAGAAGCCTGCAACCAGCAAAGGCAAGG	T10A F
KM4	CCTTGCCTTTGCTGGTTGCAGGCTTCTTGCCCTTACTTGTAGCCTTGGTAGTGGATTTCTTGGGTGCCATGTTGCTTTAGTTG	T10A R
KM5	CAACTAAAGCAACATGGCACCCAAGAAATCCACTACCAAGGAAACAAGTAAGGGCAAGAAGCCTGCAACCAGCAAAGGCAAGG	T10E F
KM6	CCTTGCCTTTGCTGGTTGCAGGCTTCTTGCCCTTACTTGTTTCCTTGGTAGTGGATTTCTTGGGTGCCATGTTGCTTTAGTTG	T10E R
KM196	TTGGCAGCGAGGGAAGCAATTATAATACAACTAAAGCAACGATTGTACTGAGAGTGCACC	HHO1::Leu2 F
KM197	TTGCTATCACCATTGACATTCTCGTTTGGATATTCACTTTACTTCTAGTATATCCACATA	HHO1::Leu2 R
KM215	AGAGGTCAGTCCAAAACCCAAGCAAGCCGCCACTTCTGTGTTCGTACGCTGCAGGTCGAC	Amplify GSKU @ HHO1 S141 F
KM216	CTAGCTTCGTGGATGCGGCTTTCGCTTTTGATGCGGTTGCCCGCGCGTTGGCCGATTCATT	Amplify GSKU @ HHO1 S141 R
KM217	AGAGGTCAGTCCAAAACCCAAGCAAGCCGCCACTTCTGTGGCTGCAACCGCATCAAAAGCGAAAGCCGCATCCACGAAGCTAG	HHO1 S141A F
KM218	CTAGCTTCGTGGATGCGGCTTTCGCTTTTGATGCGGTTGCAGCCACAGAAGTGGCGGCTTGCTTGGGTTTTGGACTGACCTCT	HHO1 S141A R
KM219	AGAGGTCAGTCCAAAACCCAAGCAAGCCGCCACTTCTGTGGAAGCAACCGCATCAAAAGCGAAAGCCGCATCCACGAAGCTAG	HHO1 S141E F
KM220	CTAGCTTCGTGGATGCGGCTTTCGCTTTTGATGCGGTTGCTTCCACAGAAGTGGCGGCTTGCTTGGGTTTTGGACTGACCTCT	HHO1 S141E R

KM230	GTCTAAGAGAGTTGGAAAAGGGTTTTGATGAAGCT GTCGCTTCGTACGCTGCAGGTCGAC	Amplify GSKU @ TRP5 F
KM231	AGAATACAAGGATTTGAAGTCTTCCCAGAATGTGG GATCGTAGGGATAACAGGGTAATCCGCGCGTTGGC CGATTCAT	Amplify GSKU @ TRP5 R
KM167F	AGAATACAAGGATTTGAAGTCTTCCCAGAATGTGG GATCGGCGACAGCTTCATCAAAACCCTTTTCCAACT CTCTTAGAC	TRP5 repair F
KM167R	GTCTAAGAGAGTTGGAAAAGGGTTTTGATGAAGCT GTCGCCGATCCCACATTCTGGGAAGACTTCAAATCC TTGTATTCT	TRP5 repair R
KM246	GAAAAAAAAATCGCCTACTGTTACCGCCAAGAAGG CCTCTTAGGGATAACAGGGTAATTTCGTACGCTGCA GGTCGAC	Amplify GSKU @ HHO1 S174 I-SceI site F
KM247	GCATGCTTTTAAGGATCATTTCCTTGTAGGTCAATG AAGACCGCGCGTTGGCCGATTCAT	Amplify GSKU @ HHO1 S174 R
KM248	AAAAAAATCGCCTACTGTTACCGCCAAGAAGGCCT CTGCTCCTTCTTCATTGACCTACAAGGAAATGATCC TTAAAAGCATGC	HHO1-S174A F
KM249	GCATGCTTTTAAGGATCATTTCCTTGTAGGTCAATG AAGAAGGAGCAGAGGCCTTCTTGGCGGTAACAGTA GGCGATTTTTTT	HHO1-S174A R
KM250	AAAAAAAAATCGCCTACTGTTACCGCCAAGAAGGC CTCTGAACCTTCTTCATTGACCTACAAGGAAATGAT CCTTAAAAGCAT	HHO1-S174E F
KM251	ATGCTTTTAAGGATCATTTCCTTGTAGGTCAATGAA GAAGGTTCAGAGGCCTTCTTGGCGGTAACAGTAGG CGATTTTTTTT	HHO1-S174E R
KM275	GGGAGTTGATCATTGAAGGGCTCACGGCTTTGAAG GAACGTAAGGGATCCTTCGTACGCTGCAGGTCGAC	HHO1-GSKU F for S65A/E mutant
KM276	GATCCGACGATCGGGTAGTTTTCCTTGATAAACTTC TTGAGTGCCGGACGCCGCGCGTTGGCCGATTCATT	HHO1-GSKU R for S65A/E mutant
KM277	TTGATCATTGAAGGGCTCACGGCTTTGAAGGAACGT AAGGGATCCGCTCGTCCGGCACTCAAGAAGTTTATC AAGGAAAACTACCCGATCGTC	HHO1-S65A F

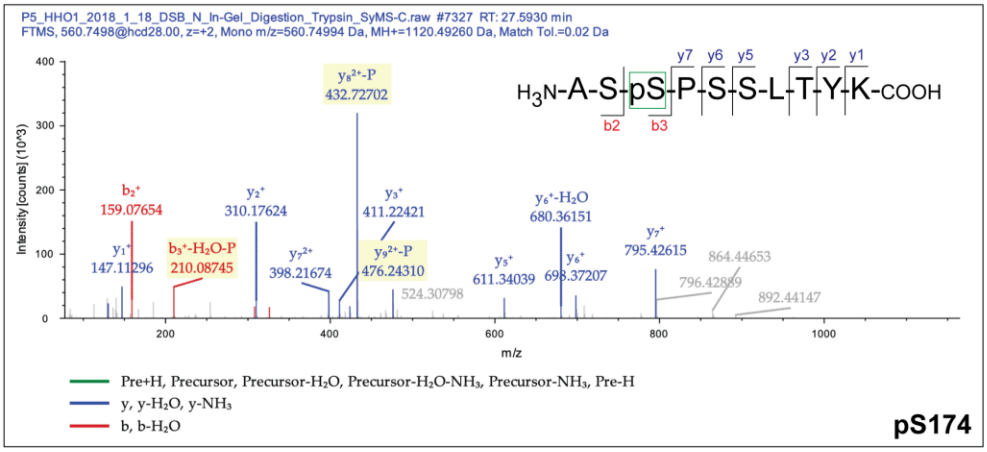
KM278	GACGATCGGGTAGTTTTCCTTGATAAACTTCTTGAG TGCCGGACGAGCGGATCCCTTACGTTCCTTCAAAGC CGTGAGCCCTTCAATGATCAA	HHO1-S65A R
KM279	TTGATCATTGAAGGGCTCACGGCTTTGAAGGAACGT AAGGGATCCGAACGTCCGGCACTCAAGAAGTTTAT CAAGGAAAACTACCCGATCGTC	HHO1-S65E F
KM280	GACGATCGGGTAGTTTTCCTTGATAAACTTCTTGAG TGCCGGACGTTCGGATCCCTTACGTTCCTTCAAAGC CGTGAGCCCTTCAATGATCAA	HHO1-S65E R
KM287	TGTAAGTCACCAATGCACTCAACGATTAGCGACCA GCCGGAATGCTTGGCTTCGTACGCTGCAGGTCGAC	His3-GSKU F
KM287	TTAACGTCCACACAGGTATAGGGTTTCTGGACCATA TGATACATGCTCTG**T**AGGGATAACAGGGTAATCCG CGCGTTGGCCGATTCAT	His3-GSKU ISceI R
KM288	CAATGCACTCAACGATTAGCGACCAGCCGGAATGC TTGGCCAGAGCATGTATCATATGGTCCAGAAACCCT ATACCTGTG	His3 repair F
KM289	CACAGGTATAGGGTTTCTGGACCATATGATACATGC TCTGGCCAAGCATTCCGGCTGGTCGCTAATCGTTGA GTGCATTG	His3 repair R
KM297	TACCGAGCTCGAATTTTTACTAACAAATGGTATTAT TTATAACAGCTAGAAGATCGGTCTGCATT	His3::Ura3 F
KM298	GTCCACACAGGTATAGGGTTTCTGGACCATATGATA CATGCTCTGGCAATGGTTGTAATGGTTTA	His3::Ura3 R
KM299	TACCGAGCTCGAATTTTTACTAACAAATGGTATTAT TTATAACAGCGAGCTCGTTTTCGACACTGG	His3::CORE F
KM300	GTCCACACAGGTATAGGGTTTCTGGACCATATGATA CATGCTCTGTCCTTACCATTAAGTTGATC	His3::CORE R
KM301	GGTACGACGTTGCTAAGGAA	DSB efficiency F
Ura3.1	TTCAATAGCTCATCAGTCGA	DSB efficiency R
KM303	GCTGCTGGTGGTCTAGGTTC	ADH1 Control F
KM304	CGGTGGTACCGTTAGCTCTA	ADH1 Control R

A

Observed Phosphosite	Peptide Spectral Matches
Non-Phosphorylated	137
pS174	19
pS173	15

* *Data from 7 LCMS runs across 4 independent experiments*

B



C

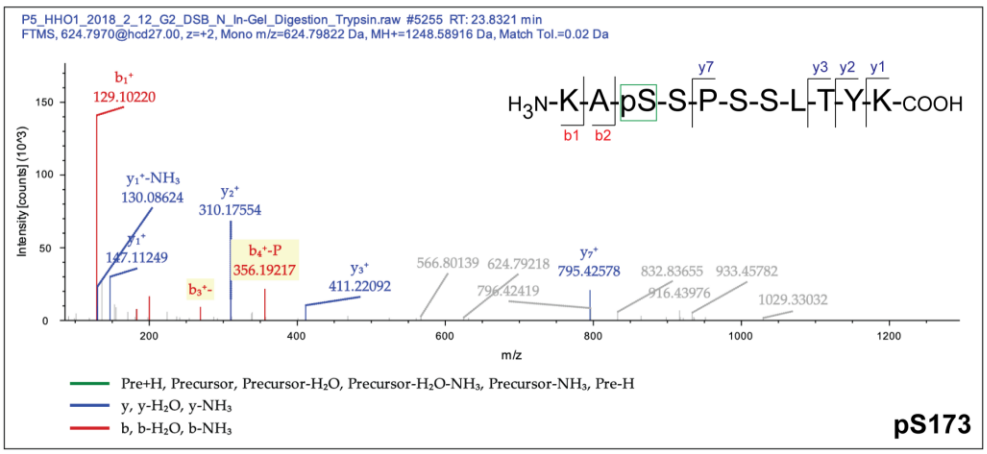


Figure A.1 Representative fragmentation mass spectra supporting evidence of Hho1 phosphorylation at both S173 and S174.

(A) Data from 7 LCMS runs (and 4 independent AP-MS experiments) were coalesced and the peptide spectral matches (PSMs) for pS173 and pS174 phosphopeptides summed, supporting evidence of phosphorylation at both sites independently but not simultaneously. A single PSM corresponds to specific spectrum generated through data-dependent fragmentation of the target peptide in an LCMS run and there are often multiple independent such spectra collected in each run – providing evidence for a given peptides sequence and modification state. (B) Representative MS/MS peptide fragmentation spectrum for a pS174 phosphopeptide. (C) Representative MS/MS peptide fragmentation spectrum for a pS173 phosphopeptide. (Insets for B,C) C-terminal (y) and N-terminal (b)

fragment ions observed and their position with respect to the Hho1 peptide sequence. Green box indicates the mapped position of phosphorylation in the observed peptides.

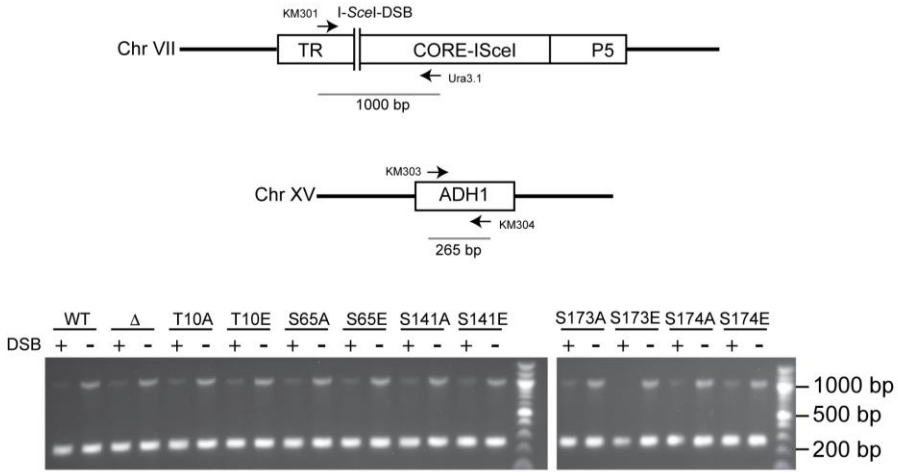


Figure A.2 I-SceI-induced DSB efficiency compared between each yeast strain.

(*Top*) Two sets of primers intended to amplify a 1,000-bp sequence surrounding the I-SceI DSB cut site in chromosome VII (KM301 & Ura3.1) and a second set intended to amplify a 265-bp sequence in the *ADH1* locus on chromosome XV (KM303 & KM304; to control for total target DNA abundance) were combined in a multiplexed PCR reaction mix distributed to an equal amount of genomic DNA extracted from each indicated strain before (-) or after (+) four hour gal-induced production of I-SceI. (*Bottom*) Representative samples from a triplicate analysis of three independent isolates for each indicated genotype showing the PCR product from each set of primers. The intensities of these bands were quantified using ImageJ and the efficiency of DSB induction measured by taking the ratio of the 1,000-bp band (including the DSB site) and the control 265-bp band.

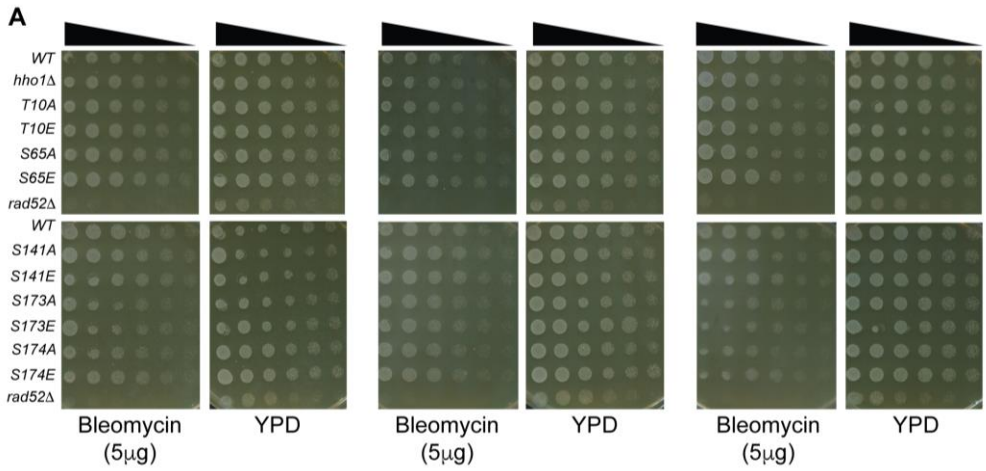


Figure A.3 Effect of Hho1 phosphosite mutants on bleomycin-induced DSB repair or on NHEJ-mediated DSB repair.

Triplicate analysis of wild type and *hho1*Δ phosphosite mutation effects on bleomycin-induced DSB induced with 5 μg bleomycin. As a control to evaluate the complete loss of DSB repair capability in cells, *rad52*Δ mutants are shown in parallel.

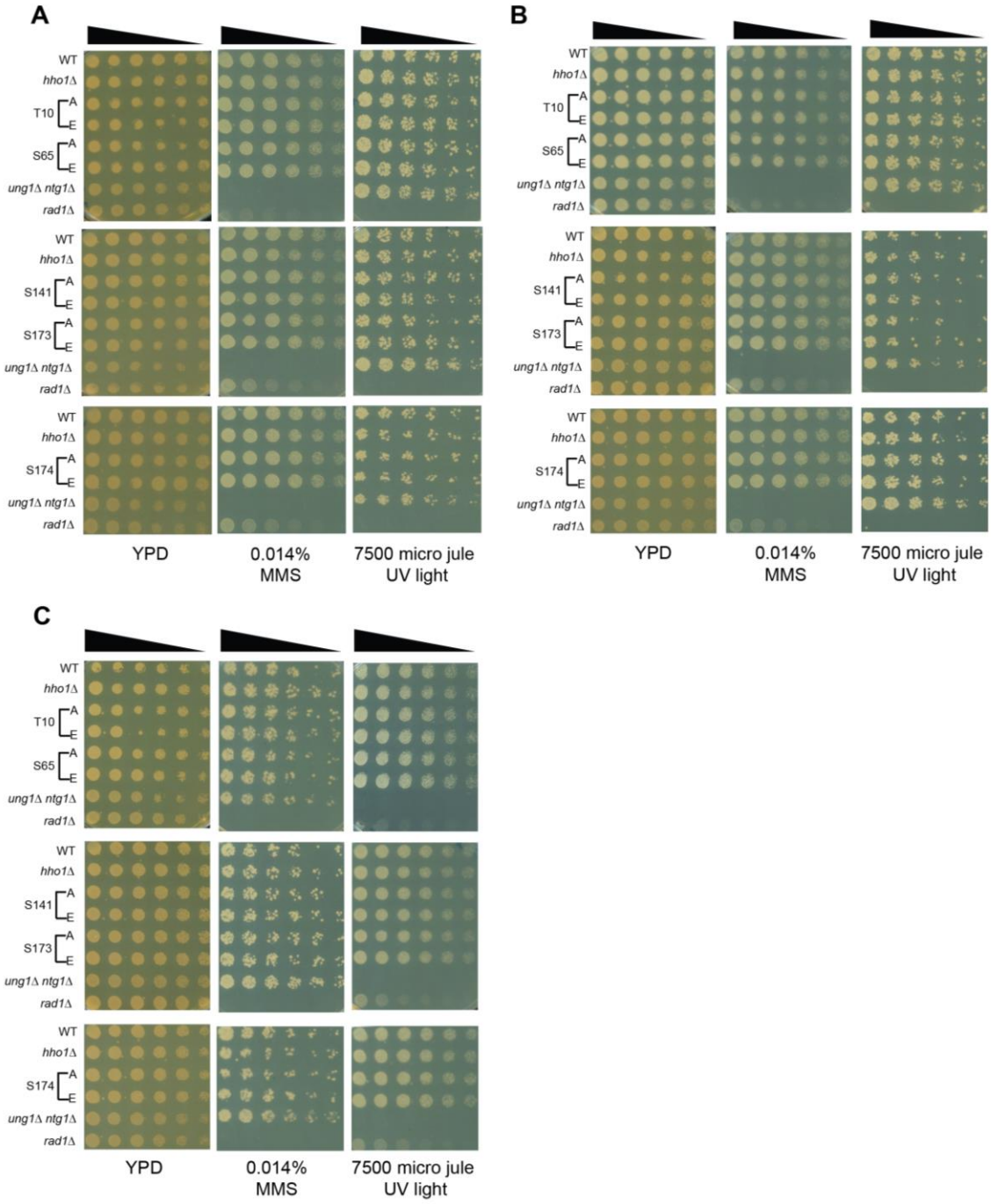


Figure A.4 Replicate spot assay results for BER and NER tests.

(A,B,C) Serial 5-fold dilution of mid log-phase cultured cells harboring wild type or the indicated phosphosite mutants were plated on non-selective media containing 0.014% MMS to induce the base excision repair pathway or were plated on non-selective media and treated with UV light in a UV cross linker, followed by outgrowth.

APPENDIX B. SUPPLEMENTARY MATERIAL FOR CHAPTER 6

Table B.1 List of strains used in this study.

Strain	Relevant genotype	Source
FRO-1073	*ho*Δ *hml*Δ::*ADE1 mata*Δ::*hisG hmr*Δ::*ADE1 ade1 leu2*::*pGAL1mhis3AI-ADE3 lys5 trp1*::*hisG ura3-52 ade3*::*GAL*::*HO*	Keskin *et al.* 2012
FRO-1075,1080	FRO-1073 *his3*::*HOcs*	Keskin *et al.* 2012
YS-164, 165	FRO-1075, 1080 (*HIS3*::*HOcs*)::*TRP1*	Keskin *et al.* 2012
YS-166, 167	YS-164, 165 *pGAL1-mhis3AI*::CORE	Keskin *et al.* 2012
YS-172, 174	YS-166, 167 *pGAL1-mhis3AI*::*HO*	Keskin *et al.* 2012
YS-278, 281	YS-172, 174 *YCLWTy2-1*::CORE	Keskin *et al.* 2012
YS-291, 292	YS-278, 281 *YCLWTy2-1*Δ	Keskin *et al.* 2012
YS-440, 441	YS-291, 292 *spt3*Δ::*kanMX4*	Keskin *et al.* 2012
YS-424, 426	YS-291, 292 *rnh1*Δ::*NAT rnh201*Δ::*kanMX4*	Keskin *et al.* 2012
YS-486, 487	YS-291, 292 *rnh1*Δ::*NAT rnh201*Δ::*hygMX4 spt3*Δ::*kanMX4*	Keskin *et al.* 2012
YS-526, 527	YS-291, 292 pGAL*1*Δ::*KlURA3*	Keskin *et al.* 2012
YS-528, 529	YS-291, 292 *rnh1*Δ::*NAT rnh201*Δ::*hygMX4 pGAL1* Δ::*KlURA3*	Keskin *et al.* 2012
YS-530, 531	YS-291, 292 *rnh1*Δ::*NAT rnh201*Δ::*hygMX4 spt3*Δ::*kanMX4 pGAL1*Δ::*KlURA3*	Keskin *et al.* 2012
CM-278, 279	YS-526, 527 (*pGAL1*Δ::*KlURA3)*::*pTEF*	this study
CM-288, 290	CM-278, 289 *spt3*Δ::*NAT*	this study
CM-280, 281, 282, 283	YS-528, 529 *rnh1*Δ::*NAT rnh201*Δ::*hygMX4* (*pGAL1* Δ::*KlURA3)*::*pTEF*	this study
CM-284, 286	YS-530, 531 *rnh1*Δ::*NAT rnh201*Δ::*hygMX4* (*pGAL1* Δ::*KlURA3)*::*pTEF*	this study
CM-320, 322	CM280, 282 *spt3*Δ::*KLURA3*	this study
CM-478, 480	CM-278, 279 *ade3*::*GAL*::*ho*Δ::*KlURA3*	this study
YS-532, 533	YS-291, 292 *ade3*::*GAL*::*ho*Δ::*KlURA3*	this study
CM-482, 483, 484, 485	CM280, 282 *rnh1*Δ::*NAT rnh201*Δ::*hygMX4 ade3*::*GAL*::*ho*Δ::*KlURA3*	this study

YS-534, 535	YS434, 426 *rnh1*Δ::*NAT rnh201*Δ::*hygMX4 ade3*::*GAL*::*ho*Δ::*KlURA3*	this study
CM-486, 488	CM320, 322 *rnh1*Δ::*NAT rnh201*Δ::*hygMX4 spt3*Δ::*kanMX4 ade3*::*GAL*::*ho*Δ::*KlURA3*	this study
YS-536, 537	YS486, 487 *rnh1*Δ::*NAT rnh201*Δ::*hygMX4 spt3*Δ::*kanMX4 ade3*::*GAL*::*ho*Δ::*KlURA3*	this study
DG-2204 *(S.P.)*	*MAT***a***his3*-Δ*200hisG trp1 ura3* Ty-less Gal^+ Spo^+	Garfinkel *et al. Genetics*, 2005
HK-692, 696 *(S.P.)*	DG-2204 *leu2*::p*GAL1-ISce-I*	this study
HK-703, 705 *(S.P.)*	HK-692, 696 *rnh201*Δ::*hygMX4*	this study
HK-805, 807 *(S.P.)*	HK-692, 696 *rnh1*Δ:: *kanMX4*	this study
HK-809, 812 *(S.P.)*	HK-703, 705 *rnh1*Δ:: *kanMX4*	this study
CM-1070	CM-278 *can1*Δ:: CORE	this study
CM-1093, 1095	CM-1070 (*can1*Δ:: CORE)::pGAL1	this study
CM-1099, 1100	CM-1070 (*can1*Δ:: CORE)::pGAL-Ty1	this study
CM-1150	CM-1095 *rnh201*Δ::*hygMX4*	this study
CM-1167, 1169	CM-1148, 1150 *rnh1*Δ::*NAT*	this study
CM-1152, 1154	CM-1099, 1100 *rnh201*Δ::*hygMX4*	this study
CM-1171, 1173, 1174	CM-1152, 1154 *rnh1*Δ::*NAT*	this study
FRO-767	*ho*Δ *hml*Δ::*ADE1 mata*Δ::*hisG hmr*Δ::*ADE1 ade1 lys5 trp1*::*hisG ura3-52 ade3*::*GAL*::*HO*	Storici *et a*l. *Nature*, 2007
KK-140	FRO-767 *(leu2::HOcs)::LEU2*	this study
HK-612, 613	KK-140 *leu2*::GSKU (*Gal1-I-SecI kanMX4 KlURA3*)	this study
HK-648	HK-612 (*leu2*::GSKU):: *Gal1-his3-AI-HO*	this study
HK-654	HK-648 AI::CORE	this study
HK-740	HK-654 (AI::CORE)::AI-I-SecI Site	this study
HK-687, 688	HK-613 *kanMX4 klura3::ADE3*	this study
HK-699, 701	HK-687, 688 *his3::TRP1*	this study
HK-791, 793	HK-699, 701 *rnh1*Δ:: *kanMX4*	this study
HK-795, 797	HK-699, 701 *rnh201*Δ::*hygMX4*	this study
HK-815, 817	HK-795, 797 *rnh1*Δ:: *kanMX4*	this study
HK-818, 820	HK-699, 701 *spt3*Δ::*NAT*	this study
HK-822, 824	HK-791,793 *spt3*Δ::*NAT*	this study
HK-826, 828	HK-795, 797 *spt3*Δ::*NAT*	this study
HK-830, 837	HK-815, 817 *spt3*Δ::*NAT*	this study
CM-1344, 1345	CM-1093, 1095 *rad1*Δ:: *kanMX4*	this study
CM-1341, 1343	CM-1099, 1100 *rad1*Δ:: *kanMX4*	this study

CM-1383, 1386	CM-1167, 1169 *rad1*Δ:: *kanMX4*	this study
CM-1379, 1381	CM-1171, 1173 *rad1*Δ:: *kanMX4*	this study
CM-452, 454	CM-278, 279 *rad10*Δ:: *kanMX4*	this study
CM-464, 466	CM-278, 279 *rad1*Δ:: *kanMX4*	this study
CM-456, 458	CM-280, 282 *rad10*Δ:: *kanMX4*	this study
CM-468, 470	CM-280, 282 *rad1*Δ:: *kanMX4*	this study
CM-460, 464	CM-320, 322 *rad10*Δ:: *kanMX4*	this study
CM-472, 474	CM-320, 322 *rad1*Δ:: *kanMX4*	this study
CM-1283, 1285	CM-1093, 1095 *ade3*::*GAL*::*ho*Δ::*kanMX4*	this study
CM-1288, 1290	CM-1099, 1100 *ade3*::*GAL*::*ho*Δ::*kanMX4*	this study
CM-1207, 1209	CM-1167, 1169 *ade3*::*GAL*::*ho*Δ::*kanMX4*	this study
CM-1211, 1213	CM-1171, 1173 *ade3*::*GAL*::*ho*Δ::*kanMX4*	this study
CM-602, 604	CM-278, 279 *msh2*Δ:: *kanMX4*	this study
CM-594, 596	CM-278, 279 *msh3*Δ:: *kanMX4*	this study
CM-620, 621	CM-278, 279 *msh6*Δ:: *kanMX4*	this study
CM-590, 591, 592, 593	CM-280, 282 *msh2*Δ:: *kanMX4*	this study
CM-598, 599, 600, 601	CM-280, 282 *msh3*Δ:: *kanMX4*	this study
CM-622, 623, 624, 625	CM-280, 282 *msh6*Δ:: *kanMX4*	this study
CM-290, 292, 293	CM-278, 279 *dnl4*Δ:: *kanMX4*	this study
CM-380, 382	CM-278, 279 *ku70*Δ:: *kanMX4*	this study
CM-294, 296	CM-280, 282 *dnl4*Δ:: *kanMX4*	this study
CM-383, 384, 386	CM-280, 281, 282, 283 *ku70*Δ:: *kanMX4*	this study
CM-343, 346	CM-278, 279 *rad59*Δ:: *kanMX4*	this study
CM-354, 356	CM-278, 279 *rad52*Δ:: *KlURA3*	this study
CM-366, 368	CM-278, 279 *rad51*Δ:: *kanMX4*	this study
CM-347, 348, 349	CM-280, 282 *rad59*Δ:: *kanMX4*	this study
CM-358, 360	CM-280, 282 *rad52*Δ:: *KlURA3*	this study
CM-370, 371, 373	CM-280, 282 *rad51*Δ:: *kanMX4*	this study
CM-350, 352	CM-320, 322 *rad59*Δ:: *kanMX4*	this study
CM-375, 377	CM-320, 322 *rad51*Δ:: *kanMX4*	this study
CM-362, 363, 364, 365	CM-284, 286 *rad52*Δ:: *KlURA3*	this study
CM-616, 617, 618, 619	CM-483, 485 *rad52*Δ:: *KlURA3*	this study
CM-1352	CM-279 *rad50*Δ:: CORE	this study
CM-1370	CM-1352 (*rad50*Δ:: CORE):: R540H T853I	this study
CM-1372	CM-1352 (*rad50*Δ:: CORE):: R540H T853I D575G	this study
CM-1389	CM-1370 *rnh201*Δ::*hygMX4*	this study
CM-1400, 1401	CM-1389 *rnh1*Δ::*natMX4*	this study
CM-1391	CM-1372 *rnh201*Δ::*hygMX4*	this study
CM-1398, 1399	CM-1391 *rnh201*Δ::*hygMX4*	this study
CM-839, 840	CM-280, 282 *rev3*Δ:: *kanMX4*	this study

CM-842, 844	CM-320, 322 *rev3*Δ:: *kanMX4*	this study
CM-903, 905	CM-280, 282 *pol32*Δ:: *kanMX4*	this study
CM-1465, 1467	CM-903, 905 *rad1*Δ:: *KlURA3*	this study
CM-573, 573	CM-280, 282 *pol4*Δ:: *kanMX4*	this study
CM-1462, 1464	CM-573, 573 *rad1*Δ:: *KlURA3*	this study
CM-1473, 1475	CM-839, 840 *rad1*Δ:: *KlURA3*	this study
CM-851, 852	CM-280, 282 *rev1*Δ:: *kanMX4*	this study
CM-1470, 1472	CM-851, 852 *rad1*Δ:: *KlURA3*	this study
CM-951, 953	CM-280, 282 *rad5*Δ:: *kanMX4*	this study
CM-1477, 1479	CM-951, 953 *rad1*Δ:: *KlURA3*	this study
CM-444, 446	CM-280, 282 *rad30*Δ:: *kanMX4*	this study
CM-1458, 1459, 1460, 1461	CM-444, 446 *rad1*Δ:: *KlURA3*	this study
CM-1165, 1166	CM-280, 282 *rev3*Δ:: CORE	this study
CM-1197, 1199	CM-1165, 1166 (*rev3*Δ:: CORE):: *rev3* L979F	this study
CM-1483, 1487	CM-1173, 1175 (*can1*Δ:: CORE)::pGAL-REV3	this study
CM-1499, 1500	CM-1483, 1487 *ho*Δ::*kanMX4*	this study
CM-1495, 1497	CM-1167, 1169 *ho*Δ::*kanMX4*	this study
CM-510, 512	CM-278, 279 2-micron plasmid	this study
CM-514, 516	CM-280, 282 2-micron plasmid	this study
CM-1298, 1299	CM-280, 282 *pms1*::*kanMX4*	this study
CM-1301, 1303	CM-280, 282 *mlh1*::*kanMX4*	this study

Table B.2 List of oligonucleotides used in this study.

Name	Length, nt	Sequence (5'→3')	Purpose
211	48	GAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAAT	D-loop and R-loop Assay
501	48	GAAGCAUUUAUCAGGGUUAUUGUCUCAUGAGCGGAUACAUAUUUGAAU (RNA)	D-loop and R-loop Assay
ACT1Q.F	20	TGTCACCAACTGGGACGATA	qRT-PCR
ACT1Q.R	20	GGCTTGGATGGAAACGTAGA	qRT-PCR
HIS3Q.1	20	CAAGTTCGACAACTGCGTAC	qRT-PCR

HIS3Q.3	20	CACATAGACGACCATCACAC	qRT-PCR
HIS3.F	80	ACCAATGCACTCAACGATTAGCGACCAGCCGGAATGCTTGGCCAGAGCATGTATCATATGGTCCAGAAACCCTATACCTG	Oligo Assay
HIS3.R	80	CAGGTATAGGGTTTCTGGACCATATGATACATGCTCTGGCCAAGCATTCCGGCTGGTCGCTAATCGTTGAGTGCATTGGT	Oligo Assay
HIS3.205F	20	GACCGAGAGCAATCCCGCAG	PCR Product Repair Assay
HIS3.205R	20	ATCTTCCCAGAAAAAGAGGC	PCR Product Repair Assay

Table B.3. *P*-value of comparisons of His$^+$ repair frequencies in the *pGAL1 vs pTEF* system.

	WT (***pGAL1)***	*spt3* (***pGAL1)***	*rnh1 rnh201* ***(pGAL1)***	*rnh1 rnh201 spt3* ***(pGAL1)***	WT (***pTEF)***	*spt3* (***pTEF)***	*rnh1 rnh201* ***(pTEF)***	*rnh1 rnh201 spt3* ***(pTEF)***
WT (***pGAL1)***		0.0022**	0.0022**	0.0022**	0.0022**	0.0022**	0.0022**	0.0022**
spt3 (***pGAL1)***	0.0022**		0.0022**	0.0022**	0.0152*	>0.9999	0.0022**	0.0022**
rnh1 rnh201 ***(pGAL1)***	0.0022**	0.0022**		0.0022**	0.0022**	0.0022**	0.3095	0.0022**
rnh1 rnh201 spt3 ***(pGAL1)***	0.0022**	0.0022**	0.0022**		0.0022**	0.0022**	0.0022**	0.0022**
WT (***pTEF)***	0.0022**	0.0152*	0.0022**	0.0022**		0.0152*	0.0022**	0.0022**
spt3 (***pTEF)***	0.0022**	>0.9999	0.0022**	0.0022**	0.0152*		0.0022**	0.0022**
rnh1 rnh201 ***(pTEF)***	0.0022**	0.0022**	0.3095	0.0022**	0.0022**	0.0022**		0.0022**
rnh1 rnh201 spt3 ***(pTEF)***	0.0022**	0.0022**	0.0022**	0.0022**	0.0022**	0.0022**	0.0022**	

Table B.4 *P*-value of comparisons of His$^+$ repair frequencies of recombination mutants.

	WT	*rnh1 rnh201*	*rnh1 rnh201 spt3*
rad59	0.0025**	0.0931	0.4567
rad52	0.0007***	0.0022**	0.0022**
rad51	0.0290*	0.1320	0.6991

Table B.5 *P*-value of comparisons of His$^+$ repair frequencies of NHEJ mutants.

	WT	*rnh1 rnh201*
ku70	0.4903	0.0022**
dnl4	0.0721	0.0022**

Table B.6 *P*-value of comparisons of His$^+$ repair frequencies of clippase mutants.

	WT	*rnh1 rnh201*	*rnh1 rnh201 spt3*
rad1	<0.0001***	0.0022**	0.0087**
rad10	<0.0001***	0.0043**	0.0043**

Table B.7 *P*-value of comparisons of His$^+$ repair frequencies of mismatch repair mutants.

	WT	*rnh1 rnh201*
msh2	0.0029**	<0.0001***
msh3	0.0029**	<0.0001***
msh6	0.8735	<0.0001***

Table B.8 *P*-value of comparisons of His$^+$ repair frequencies in A. *S. cerevisiae (Sc)* and B. *S. paradoxus (Sp).*

A.

	WT (Sc)	*spt3 (Sc)*	*rnh1 (Sc)*	*spt3 rnh1 (Sc)*	*rnh201 (Sc)*	*spt3 rnh201 (Sc)*	*rnh1 rnh201 (Sc)*	*rnh1 rnh201 spt3 (Sc)*
WT (Sc)		0.0001***	0.0036**	0.0001***	<0.0001****	0.0001***	<0.0001****	0.0001***
spt3 (Sc)	0.0001***		0.0001***	>0.9999	0.0001***	>0.9999	0.0001***	0.0022**
rnh1 (Sc)	0.0036**	0.0001***		0.0001***	>0.9999	0.0001***	<0.0001****	0.0001***
spt3 rnh1 (Sc)	0.0001***	>0.9999	0.0001***		0.0001***	>0.9999	0.0001***	0.0022**
rnh201 (Sc)	<0.0001****	0.0001***	>0.9999	0.0001***		0.0001***	<0.0001****	0.0001***
spt3 rnh201 (Sc)	0.0001***	>0.9999	0.0001***	>0.9999	0.0001***		0.0001***	0.0022**
rnh1 rnh201 (Sc)	<0.0001****	0.0001***	<0.0001****	0.0001***	<0.0001****	0.0001***		0.1025
rnh1 rnh201 spt3 (Sc)	0.0001***	0.0022**	0.0001***	0.0022**	0.0001***	0.0022**	0.1025	

B.

	WT (*Sp*)	*rnh1 (Sp)*	*rnh201 (Sp)*	*rnh1 rnh201 (Sp)*
WT *(Sp)*		>0.9999	>0.9999	<0.0001***
rnh1 (Sp)	>0.9999		>0.9999	<0.0001***
rnh201 (Sp)	>0.9999	>0.9999		<0.0001***
rnh1 rnh201 (Sp)	<0.0001***	<0.0001***	<0.0001***	

Table B.9 *P*-value of comparisons of His$^+$ repair frequencies with Ty overexpression.

	WT pGAL1	*WT pGAL1-Ty*	*rad1 pGAL1*	*rad1 pGAL1-Ty*	*rnh1 rnh201 pGAL1*	*rnh1 rnh201 pGAL1-Ty*	*rnh1 rnh201 rad1 pGAL1*	*rnh1 rnh201 rad1 pGAL1-Ty*
WT pGAL1		0.0022**	0.0022**	0.6775	0.0022**	0.0022**	0.0022**	0.0022**
WT pGAL1-Ty	0.0022**		0.0022**	0.0022**	0.0022**	0.0022**	0.0022**	0.0022**
rad1 pGAL1	0.0022**	0.0022**		0.0022**	0.0022**	0.0022**	0.0022**	0.0022**
rad1 pGAL1-Ty	0.6775	0.0022**	0.0022**		0.0022**	0.0022**	0.0022**	0.0022**
rnh1 rnh201 pGAL1	0.0022**	0.0022**	0.0022**	0.0022**		0.0022**	0.3095	0.5584
rnh1 rnh201 pGAL1-Ty	0.0022**	0.0022**	0.0022**	0.0022**	0.0022**		0.0043**	0.0022**
rnh1 rnh201 rad1 pGAL1	0.0022**	0.0022**	0.0022**	0.0022**	0.3095	0.0043**		0.0606
rnh1 rnh201 rad1 pGAL1-Ty	0.0022**	0.0022**	0.0022**	0.0022**	0.5584	0.0022**	0.0606	

Table B.10 *P*-value of comparisons of His$^+$ repair frequencies of non-essential polymerases.

	rnh1 rnh201	*rnh1 rnh201 rad1*
pol4	<0.0001***	0.0001***
rad30	<0.0001***	0.1432
pol32	<0.0001***	0.2129
rev1	<0.0001***	0.0013**
rev3	<0.0001***	0.0068**
rad5	<0.0001***	0.0045**

Table B.11 *P*-value of comparisons of His$^+$ repair frequencies in the absence of DSB.

	rnh1 rnh201	*rnh1 rnh201 rad1*	*rnh1 rnh201 rad51*	*rnh1 rnh201 rad59*	*rnh1 rnh201 spt3*	*rnh1 rnh201 spt3 rad52*	*rnh1 rnh201 ho*	*rnh1 rnh201 ho rad52*
rnh1 rnh201		0.0318*	0.0238*	0.1255	0.1095	0.0095**	0.0318*	0.0389
rnh1 rnh201 rad1	0.0318*		0.8619	0.0095**	0.6857	0.6857	0.2902	0.6167
rnh1 rnh201 rad51	0.0238*	0.8619		0.0043**	0.5619	0.8619	0.3215	0.7901
rnh1 rnh201 rad59	0.1255	0.0095**	0.0043**		0.0952	0.0095**	0.0023**	0.0058**
rnh1 rnh201 spt3	0.1095	0.6857	0.5619	0.0952		0.4857	0.7505	0.3847
rnh1 rnh201 spt3 rad52	0.0095**	0.6857	0.8619	0.0095**	0.4857		0.1218	0.8198
rnh1 rnh201 ho	0.0318*	0.2902	0.3215	0.0023**	0.7505	0.1218		0.1596
rnh1 rnh201 ho rad52	0.0389	0.6167	0.7901	0.0058**	0.3847	0.8198	0.1596	

Table B.12 *P*-value of comparisons of His$^+$ repair frequencies in the absence of DSB.

	rnh1 rnh201	*rnh1 rnh201 rad1*	*rnh1 rnh201 rad10*	*rnh1 rnh201 msh2*	*rnh1 rnh201 msh3*	*rnh1 rnh201 pms1*	*rnh1 rnh201 mlh1*
rnh1 rnh201		0.2121	0.5209	0.5989	0.0780	0.8615	0.9527
rnh1 rnh201 rad1	0.2121		0.8857	0.3429	0.2000	0.4857	0.6857
rnh1 rnh201 rad10	0.5209	0.8857		0.6857	0.4857	>0.9999	0.6857
rnh1 rnh201 msh2	0.5989	0.3429	0.6857		0.1143	>0.9999	0.6857
rnh1 rnh201 msh3	0.0780	0.2000	0.4857	0.1143		0.2000	0.2000
rnh1 rnh201 pms1	0.8615	0.4857	>0.9999	>0.9999	0.2000		0.8857
rnh1 rnh201 mlh1	0.9527	0.6857	0.6857	0.6857	0.2000	0.8857	

Table B.13 *P*-value of comparisons of His$^+$ repair frequencies with Ty overexpression in *ho* mutants.

	WT pGAL1	*WT pGAL1-Ty*	*ho pGAL1*	*ho pGAL1-Ty*	*rnh1 rnh201 pGAL1*	*rnh1 rnh201 pGAL1-Ty*	*rnh1 rnh201 ho pGAL1*	*rnh1 rnh201 ho pGAL1-Ty*
WT pGAL1		0.0022**	0.0022**	0.0022**	0.0022**	0.0022**	0.0022**	0.0022**
WT pGAL1-Ty	0.0022**		0.0022**	0.0022**	0.0022**	0.0022**	0.8182	0.0043**
ho pGAL1	0.0022**	0.0022**		>0.9999	0.0022**	0.0022**	0.0022**	0.0022**
ho pGAL1-Ty	0.0022**	0.0022**	>0.9999		0.0022**	0.0022**	0.0022**	0.0022**
rnh1 rnh201 pGAL1	0.0022**	0.0022**	0.0022**	0.0022**		0.0043**	0.0022**	0.0022**
rnh1 rnh201 pGAL1-Ty	0.0022**	0.0022**	0.0022**	0.0022**	0.0043**		0.0022**	0.0022**
rnh1 rnh201 ho pGAL1	0.0022**	0.8182	0.0022**	0.0022**	0.0022**	0.0022**		0.0022**
rnh1 rnh201 ho pGAL1-Ty	0.0022**	0.0043**	0.0022**	0.0022**	0.0022**	0.0022**	0.0022**	

Table B.14 *P*-value of comparisons of His$^+$ repair frequencies of non-essential polymerases in the absence of DSB.

	rnh1 rnh201	*rnh1 rnh201 pol4*	*rnh1 rnh201 rad30*	*rnh1 rnh201 pol32*	*rnh1 rnh201 rev1*	*rnh1 rnh201 rev3*	*rnh1 rnh201 rev3 L979F*	*rnh1 rnh201 rad5*
rnh1 rnh201		0.6828	0.8081	0.1535	0.0040**	0.0040**	0.0040**	0.0283*
rnh1 rnh201 pol4	0.6828		>0.9999	>0.9999	0.0286*	0.0286*	0.0286*	0.1143
rnh1 rnh201 rad30	0.8081	>0.9999		0.2000	0.0286*	0.0286*	0.0286*	0.0286*
rnh1 rnh201 pol32	0.1535	>0.9999	0.2000		0.0286*	0.0286*	0.0286*	0.0286*
rnh1 rnh201 rev1	0.0040**	0.0286*	0.0286	0.0286*		>0.9999	0.1429	0.0286*
rnh1 rnh201 rev3	0.0040**	0.0286*	0.0286*	0.0286*	>0.9999		0.1429	0.0286*
rnh1 rnh201 rev3 L979F	0.0040**	0.0286*	0.0286*	0.0286*	0.1429	0.1429		0.4857
rnh1 rnh201 rad5	0.0283*	0.1143	0.0286*	0.0286*	0.0286*	0.0286*	0.4857	

Table B.15 *P*-value of comparisons of DNA repair frequencies obtained after transformation with DNA oligonucleotides (A) or without DNA oligonucleotides (B) following DSB induction in *his3* in the *pGAL1 vs pTEF* system.

A.

	WT- pGAL	*WT- pTEF*	*rnh1 rnh201 spt3- pGAL*	*rnh1 rnh201 spt3- pTEF*
WT- pGAL		>0.9999	>0.9999	0.2000
WT- pTEF	>0.9999		0.0286*	0.0286*
rnh1 rnh201 spt3- pGAL	>0.9999	0.0286*		0.0286*
rnh1 rnh201 spt3- pTEF	0.2000	0.0286*	0.0286*	

B.

	WT- pGAL	*WT- pTEF*	*rnh1 rnh201 spt3- pGAL*	*rnh1 rnh201 spt3- pTEF*
WT- pGAL		>0.9999	0.0286*	0.0286*
WT- pTEF	>0.9999		0.0286*	0.0286*
rnh1 rnh201 spt3- pGAL	0.0286*	0.0286*		0.0286*
rnh1 rnh201 spt3- pTEF	0.0286*	0.0286*	0.0286*	

Table B.16 *P*-value of comparisons of DNA repair frequencies with *his3* DNA oligonucleotides following DSB induction for (A) *S. cerevisiae* (*Sc*) or (B) *S. paradoxus* (*Sp*).

A.

	WT (Sc)	*spt3 (Sc)*	*rnh1 (Sc)*	*spt3 rnh1 (Sc)*	*rnh201 (Sc)*	*spt3 rnh201 (Sc)*	*rnh1 rnh201 (Sc)*	*rnh1 rnh201 spt3 (Sc)*
WT (Sc)		0.1143	>0.9999	0.0286*	0.3429	0.2000	0.3429	0.1143
spt3 (Sc)	0.1143		0.0286*	0.1143	0.0286*	>0.9999	0.0286*	0.8857
rnh1 (Sc)	>0.9999	0.0286*		0.0286*	0.3429	0.2000	0.0571	0.1143
spt3 rnh1 (Sc)	0.0286*	0.1143	0.0286*		0.0286*	0.8857	0.0286*	0.3429
rnh201 (Sc)	0.3429	0.0286*	0.3429	0.0286*		0.6857	0.0286*	0.6857
spt3 rnh201 (Sc)	0.2000	>0.9999	0.2000	0.8857	0.6857		0.0286*	0.8857
rnh1 rnh201 (Sc)	0.3429	0.0286*	0.0571	0.0286*	0.0286*	0.0286*		0.0286*
rnh1 rnh201 spt3 (Sc)	0.1143	0.8857	0.1143	0.3429	0.6857	0.8857	0.0286*	

B.

	WT (Sp)	*rnh1 (Sp)*	*rnh201 (Sp)*	*rnh1 rnh201 (Sp)*
WT (Sp)		0.3429	0.0286*	0.0286*
rnh1 (Sp)	0.3429		0.0286*	0.0286*
rnh201 (Sp)	0.0286*	0.0286*		0.0286*
rnh1 rnh201 (Sp)	0.0286*	0.0286*	0.0286*	

Table B.17 *P*-value of comparisons of His$^+$ repair frequencies of *pGAL1* system (A) and *pTEF* system (B) overexpressing Ty transposon.

A.

	WT – pBDG102	*WT – pGTyClaI*	*rnh1 rnh201 - pBDG102*	*rnh1 rnh201 - pGTyClaI*	*rnh1 rnh201 spt3 - pBDG102*	*rnh1 rnh201 spt3 - pGTyClaI*
WT – pBDG102		0.0022**	0.0022**	0.0022**	0.0022**	0.0022**
WT – pGTyClaI	0.0022**		0.0022**	0.0022**	0.0022**	0.0022**
rnh1 rnh201 - pBDG102	0.0022**	0.0022**		0.0022**	0.0022**	0.1797
rnh1 rnh201 - pGTyClaI	0.0022**	0.0022**	0.0022**		0.0022**	0.0022**
rnh1 rnh201 spt3 - pBDG102	0.0022**	0.0022**	0.0022**	0.0022**		0.0022**
rnh1 rnh201 spt3 - pGTyClaI	0.0022**	0.0022**	0.1797	0.0022**	0.0022**	

B.

	WT – pBDG102	*WT – pGTyClaI*	*rnh1 rnh201 - pBDG102*	*rnh1 rnh201 - pGTyClaI*
WT – pBDG102		0.0022**	0.0022**	0.0022**
WT – pGTyClaI	0.0022**		0.0022**	0.2251
rnh1 rnh201 - pBDG102	0.0022**	0.0022**		0.0022**
rnh1 rnh201 - pGTyClaI	0.0022**	0.2251	0.0022**	

Table B.18 *P*-value of comparisons of DNA repair frequencies with *his3* dsDNA oligonucleotides following DSB induction for strains with and without integrated *Ty* overexpression.

	pGAL1	*pGAL-Ty*	*rnh1 rnh201 pGAL1*	*rnh1 rnh201 pGAL1-Ty*
pGAL1		0.8857	0.6857	0.3429
pGAL-Ty	0.8857		0.4857	0.1143
rnh1 rnh201 pGAL1	0.6857	0.4857		0.1143
rnh1 rnh201 pGAL1-Ty	0.3429	0.1143	0.1143	

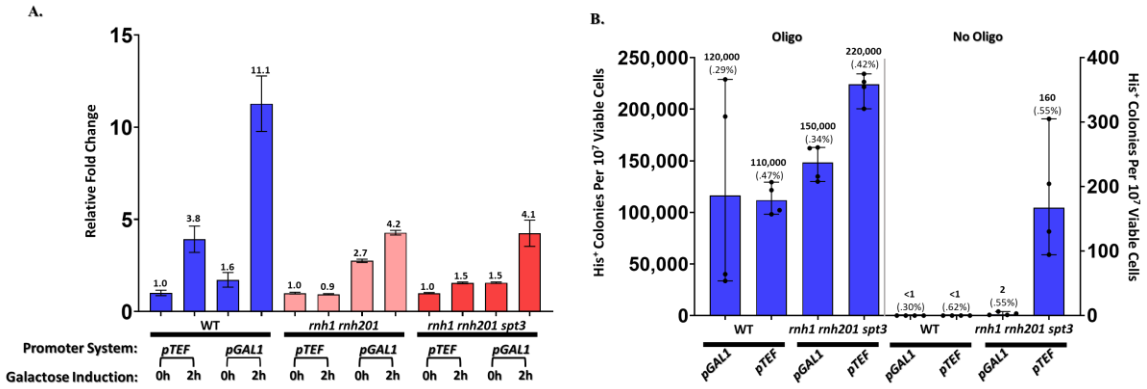


Figure B.1 Constitutive expression of donor RNA increases R-TDR.

(A) qRT-PCR of *his3* transcript RNA from wild-type, *rnh1 rnh201* and *rnh1 rnh201 spt3* mutant cells lacking the *HO* endonuclease gene and expressing the *his3* antisense RNA from either the constitutive *pTEF* or the inducible *pGAL1* promoter. Measures were taken at 0h and 2h of galactose treatment, which activates the *pGAL1* system for expression of *his3* antisense RNA. Fold change is shown relative to abundance of *his3* transcript at 0h in the *pTEF* system in each genotype (WT, *rnh1 rnh201* and *rnh1 rnh201 spt3*). (B) Frequencies of His^+ colonies formed following transformation and DSB induction with and without DNA oligos HIS3.F and HIS3.R. Cells used were wild-type or *rnh1 rnh201 spt3* mutant expressing the *his3* antisense RNA either from *pGAL1* or *pTEF*. *P* values of comparisons are shown in **Tables B.15**.

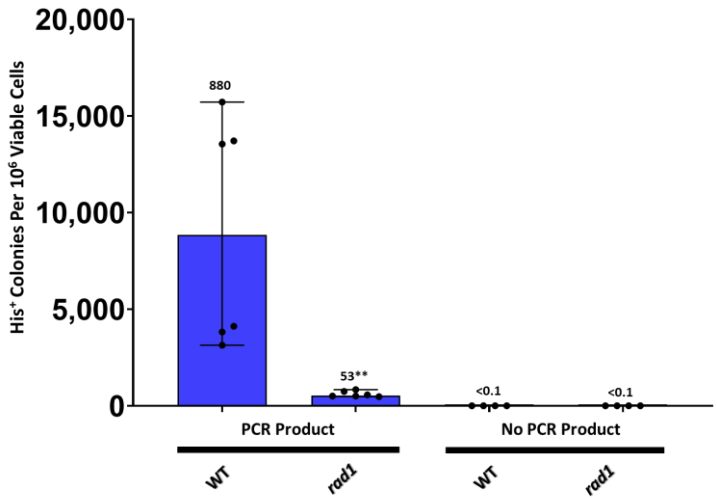


Figure B.2 Repair of *his3* DSB with PCR product in *rad1*-null cells.

Fluctuation assay showing frequency of His^+ colonies per 10^6 viable cells following DSB induction in *rad1* mutant cells transformed with PCR product. Individual frequencies are plotted. Bars represent median with 95% confidence interval. The genotype of the samples is indicated under each bar. The median is shown above each bar N=4-6. *P*-value WT-*rad1* + PCR Product (0.0022).

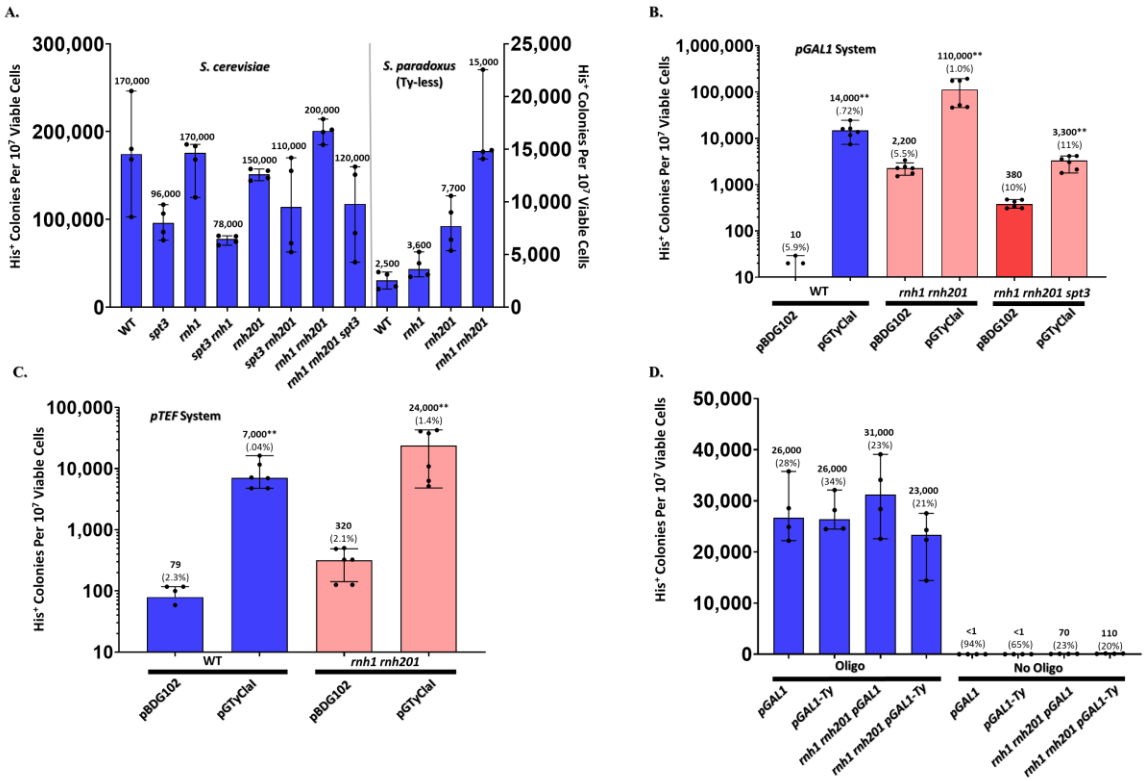


Figure B.3 Effects of Ty transposon on DSB repair.

(A) Repair frequency per 10^7 viable cells of *his3* DSB in *S. cerevisiae* and *S. paradoxus* with 1nM of ssDNA oligo. Median with 95% confidence interval is shown. N=4. *P* values of comparisons are shown in **Table B.16**. (B) Fluctuation assay showing frequency of His^+ colonies per 10^7 viable cells in *pGAL1* donor RNA system with Ty overexpression. pBDG102 represents empty plasmid. pGTyClaI represents plasmid containing *pGAL1-Ty*. Individual frequencies are plotted. Bars represent median with 95% confidence interval. The median is shown above each bar and survival shown in parentheses; N=6 *P* values of comparisons are shown in **Table B.17A**. (C) Fluctuation assay showing frequency of His^+ colonies per 10^7 viable cells in *pTEF* donor RNA system with Ty overexpression. pBDG102 represents empty plasmid. pGTyClaI represents plasmid containing *pGAL1-Ty*. Individual frequencies are plotted. Bars represent median with 95% confidence interval. The median is shown above each bar and survival shown in parentheses; N=6 *P* values of comparisons are shown in **Table B.17B**. (D) Repair frequencies by ssDNA oligos of *his3* locus following DSB induction in wild-type and *rnh1 rnh201* mutants overexpressing integrated Ty. There is no statically significant difference in the DSB repair frequency by oligos between cells in which Ty is overexpressed under *pGAL1* and cells with no Ty under *pGAL1*. *P* values of comparisons are shown in **Table B.18**.

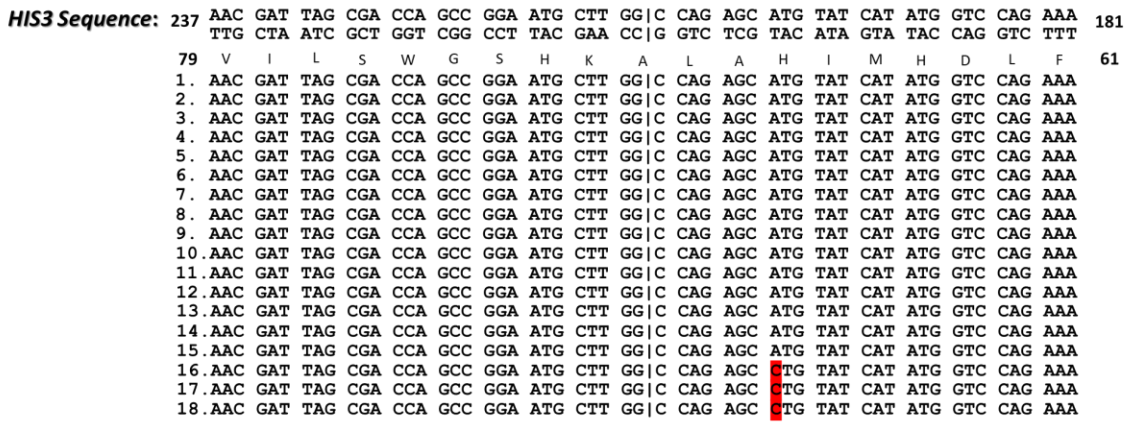


Figure B.4 Low fidelity DNA Polymerase ζ incorporates errors around the break site.

Alignment of *his3* sequences derived from 18 His^+ colony isolates of *rnh1 rnh201 rev3* L979F mutant cells. The black vertical line illustrates the site of exon-exon junction. Amino acid sequence is shown coding from right to left. Nucleotide and amino acid location are listed relative to *HIS3* start codon. Mutations found are outlined in red.

CHAPTER 8. REFERENCES

1. Desouky, O., N. Ding, and G. Zhou, *Targeted and non-targeted effects of ionizing radiation.* Journal of Radiation Research and Applied Sciences, 2015. **8**(2): p. 247-254.

2. Vignard, J., G. Mirey, and B. Salles, *Ionizing-radiation induced DNA double-strand breaks: a direct and indirect lighting up.* Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology, 2013. **108**(3): p. 362-369.

3. Chatterjee, N. and G.C. Walker, *Mechanisms of DNA damage, repair, and mutagenesis.* Environmental and molecular mutagenesis, 2017. **58**(5): p. 235-263.

4. Obe, G., C. Johannes, and D. Schulte-Frohlinde, *DNA double-strand breaks induced by sparsely ionizing radiation and endonucleases as critical lesions for cell death, chromosomal aberrations, mutations and oncogenic transformation.* Mutagenesis, 1992. **7**(1): p. 3-12.

5. Bernard, J.J., R.L. Gallo, and J. Krutmann, *Photoimmunology: how ultraviolet radiation affects the immune system.* Nat Rev Immunol, 2019. **19**(11): p. 688-701.

6. Wyatt, M.D. and D.L. Pittman, *Methylating agents and DNA repair responses: Methylated bases and sources of strand breaks.* Chem Res Toxicol, 2006. **19**(12): p. 1580-94.

7. John, K., et al., *Benzo[a]pyrene (BP) DNA adduct formation in DNA repair-deficient p53 haploinsufficient [Xpa(-/-)p53(+/-)] and wild-type mice fed BP and BP plus chlorophyllin for 28 days.* Carcinogenesis, 2012. **33**(11): p. 2236-2241.

8. McVey, M., et al., *Eukaryotic DNA Polymerases in Homologous Recombination.* Annu Rev Genet, 2016. **50**: p. 393-421.

9. Kunkel, T.A., *Balancing eukaryotic replication asymmetry with replication fidelity.* Curr Opin Chem Biol, 2011. **15**(5): p. 620-6.

10. Ganai, R.A. and E. Johansson, *DNA Replication-A Matter of Fidelity.* Mol Cell, 2016. **62**(5): p. 745-55.

11. Vaisman, A. and R. Woodgate, *Redundancy in ribonucleotide excision repair: Competition, compensation, and cooperation.* DNA Repair (Amst), 2015. **29**: p. 74-82.

12. Wang, J.C., *Cellular roles of DNA topoisomerases: a molecular perspective.* Nat Rev Mol Cell Biol, 2002. **3**(6): p. 430-40.

13. Stewart, L., et al., *A model for the mechanism of human topoisomerase I.* Science, 1998. **279**(5356): p. 1534-41.

14. Pommier, Y., et al., *Repair of topoisomerase I-mediated DNA damage.* Prog Nucleic Acid Res Mol Biol, 2006. **81**: p. 179-229.

15. Kavli, B., et al., *Uracil in DNA--general mutagen, but normal intermediate in acquired immunity.* DNA Repair (Amst), 2007. **6**(4): p. 505-16.

16. Pilzecker, B. and H. Jacobs, *Mutating for Good: DNA Damage Responses During Somatic Hypermutation.* Front Immunol, 2019. **10**: p. 438.

17. Chemudupati, M., et al., *From APOBEC to ZAP: Diverse mechanisms used by cellular restriction factors to inhibit virus infections.* Biochim Biophys Acta Mol Cell Res, 2019. **1866**(3): p. 382-394.

18. Wallace, S.S., *Base excision repair: a critical player in many games.* DNA repair, 2014. **19**: p. 14-26.

19. Chan, K., M.A. Resnick, and D.A. Gordenin, *The choice of nucleotide inserted opposite abasic sites formed within chromosomal DNA reveals the polymerase activities participating in translesion DNA synthesis.* DNA Repair (Amst), 2013. **12**(11): p. 878-89.

20. Zhao, R.Z., et al., *Mitochondrial electron transport chain, ROS generation and uncoupling (Review).* Int J Mol Med, 2019. **44**(1): p. 3-15.

21. AbdulSalam, S.F., F.S. Thowfeik, and E.J. Merino, *Excessive Reactive Oxygen Species and Exotic DNA Lesions as an Exploitable Liability.* Biochemistry, 2016. **55**(38): p. 5341-5352.

22. Prunier, A.L. and R. Leclercq, *Role of mutS and mutL genes in hypermutability and recombination in Staphylococcus aureus.* J Bacteriol, 2005. **187**(10): p. 3455-64.

23. Guarne, A., *The functions of MutL in mismatch repair: the power of multitasking.* Prog Mol Biol Transl Sci, 2012. **110**: p. 41-70.

24. Fukui, K., *DNA mismatch repair in eukaryotes and bacteria.* Journal of nucleic acids, 2010. **2010**: p. 260512.

25. Krokan, H.E. and M. Bjoras, *Base excision repair.* Cold Spring Harb Perspect Biol, 2013. **5**(4): p. a012583.

26. Meas, R., J.J. Wyrick, and M.J. Smerdon, *Nucleosomes Regulate Base Excision Repair in Chromatin.* Mutat Res, 2019. **780**: p. 29-36.

27. Schärer, O.D., *Nucleotide excision repair in eukaryotes.* Cold Spring Harbor perspectives in biology, 2013. **5**(10): p. a012609-a012609.

28. Chang, H.H.Y., et al., *Non-homologous DNA end joining and alternative pathways to double-strand break repair.* Nat Rev Mol Cell Biol, 2017. **18**(8): p. 495-506.

29. Ceccaldi, R., B. Rondinelli, and A.D. D'Andrea, *Repair Pathway Choices and Consequences at the Double-Strand Break.* Trends Cell Biol, 2016. **26**(1): p. 52-64.

30. Pertea, M., *The human transcriptome: an unfinished story.* Genes, 2012. **3**(3): p. 344-360.

31. Crossley, M.P., M. Bocek, and K.A. Cimprich, *R-Loops as Cellular Regulators and Genomic Threats.* Mol Cell, 2019. **73**(3): p. 398-411.

32. Cerritelli, S.M. and R.J. Crouch, *Ribonuclease H: the enzymes in eukaryotes.* The FEBS journal, 2009. **276**(6): p. 1494-1505.

33. Ohle, C., et al., *Transient RNA-DNA Hybrids Are Required for Efficient Double-Strand Break Repair.* Cell, 2016. **167**(4): p. 1001-1013.e7.

34. Amon, J.D. and D. Koshland, *RNase H enables efficient repair of R-loop induced DNA damage.* Elife, 2016. **5**.

35. Yasuhara, T., et al., *Human Rad52 Promotes XPG-Mediated R-loop Processing to Initiate Transcription-Associated Homologous Recombination Repair.* Cell, 2018. **175**(2): p. 558-570.e11.

36. Francia, S., et al., *Site-specific DICER and DROSHA RNA products control the DNA-damage response.* Nature, 2012. **488**(7410): p. 231-235.

37. Yu, X. and A. Gabriel, *Patching broken chromosomes with extranuclear cellular DNA.* Mol Cell, 1999. **4**(5): p. 873-81.

38. Morrish, T.A., et al., *DNA repair mediated by endonuclease-independent LINE-1 retrotransposition.* Nat Genet, 2002. **31**(2): p. 159-65.

39. Esnault, C., J. Maestre, and T. Heidmann, *Human LINE retrotransposons generate processed pseudogenes.* Nat Genet, 2000. **24**(4): p. 363-7.

40. Keskin, H., et al., *Transcript-RNA-templated DNA recombination and repair.* Nature, 2014. **515**(7527): p. 436-9.

41. Steele, E.J., *Reverse Transcriptase Mechanism of Somatic Hypermutation: 60 Years of Clonal Selection Theory.* Frontiers in immunology, 2017. **8**: p. 1611-1611.

42. Burns, J., et al., *Recurring patterns among scrambled genes in the encrypted genome of the ciliate Oxytricha trifallax.* Journal of theoretical biology, 2016. **410**: p. 171-180.

43. Fang, W. and L.F. Landweber, *RNA-mediated genome rearrangement: hypotheses and evidence.* Bioessays, 2013. **35**(2): p. 84-7.

44. Nowacki, M., et al., *RNA-mediated epigenetic programming of a genome-rearrangement pathway.* Nature, 2008. **451**(7175): p. 153-8.

45. Storici, F., et al., *RNA-templated DNA repair.* Nature, 2007. **447**(7142): p. 338-41.

46. Shen, Y., et al., *RNA-driven genetic changes in bacteria and in human cells.* Mutat Res, 2011. **717**(1-2): p. 91-8.

47. Mazina, O.M., et al., *Rad52 Inverse Strand Exchange Drives RNA-Templated DNA Double-Strand Break Repair.* Molecular cell, 2017. **67**(1): p. 19-29.e3.

48. Derr, L.K., J.N. Strathern, and D.J. Garfinkel, *RNA-mediated recombination in S. cerevisiae.* Cell, 1991. **67**(2): p. 355-64.

49. Storici, F. and M.A. Resnick, *Delitto perfetto targeted mutagenesis in yeast with oligonucleotides.* Genet Eng (N Y), 2003. **25**: p. 189-207.

50. Baltimore, D., *Retroviruses and retrotransposons: the role of reverse transcription in shaping the eukaryotic genome.* Cell, 1985. **40**(3): p. 481-2.

51. Autexier, C. and N.F. Lue, *The structure and function of telomerase reverse transcriptase.* Annu Rev Biochem, 2006. **75**: p. 493-517.

52. Luan, D.D., et al., *Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition.* Cell, 1993. **72**(4): p. 595-605.

53. Curcio, M.J. and D.J. Garfinkel, *Single-step selection for Ty1 element retrotransposition.* Proc Natl Acad Sci U S A, 1991. **88**(3): p. 936-40.

54. Derr, L.K. and J.N. Strathern, *A role for reverse transcripts in gene conversion.* Nature, 1993. **361**(6408): p. 170-3.

55. Moore, J.K. and J.E. Haber, *Capture of retrotransposon DNA at the sites of chromosomal double-strand breaks.* Nature, 1996. **383**(6601): p. 644-6.

56. Teng, S.C., B. Kim, and A. Gabriel, *Retrotransposon reverse-transcriptase-mediated repair of chromosomal breaks.* Nature, 1996. **383**(6601): p. 641-4.

57. Melamed, C., Y. Nevo, and M. Kupiec, *Involvement of cDNA in homologous recombination between Ty elements in Saccharomyces cerevisiae.* Mol Cell Biol, 1992. **12**(4): p. 1613-20.

58. Nevo-Caspi, Y. and M. Kupiec, *Transcriptional induction of Ty recombination in yeast.* Proc Natl Acad Sci U S A, 1994. **91**(26): p. 12711-5.

59. Sharon, G., T.J. Burkett, and D.J. Garfinkel, *Efficient homologous recombination of Ty1 element cDNA when integration is blocked.* Mol Cell Biol, 1994. **14**(10): p. 6540-51.

60. Nevo-Caspi, Y. and M. Kupiec, *cDNA-mediated Ty recombination can take place in the absence of plus-strand cDNA synthesis, but not in the absence of the integrase protein.* Curr Genet, 1997. **32**(1): p. 32-40.

61. Morrish, T.A., et al., *Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres.* Nature, 2007. **446**(7132): p. 208-12.

62. Storici, F., *RNA-mediated DNA modifications and RNA-templated DNA repair.* Curr Opin Mol Ther, 2008. **10**(3): p. 224-30.

63. Fink, G.R., *Pseudogenes in yeast?* Cell, 1987. **49**(1): p. 5-6.

64. Niu, D.K., W.R. Hou, and S.W. Li, *mRNA-mediated intron losses: evidence from extraordinarily large exons.* Mol Biol Evol, 2005. **22**(6): p. 1475-81.

65. Gargouri, A., *The reverse transcriptase encoded by ai1 intron is active in trans in the retro-deletion of yeast mitochondrial introns.* FEMS Yeast Res, 2005. **5**(9): p. 813-22.

66. Murakami, E., et al., *Characterization of novel reverse transcriptase and other RNA-associated catalytic activities by human DNA polymerase gamma: importance in mitochondrial DNA replication.* J Biol Chem, 2003. **278**(38): p. 36403-9.

67. Trott, D.A. and A.C. Porter, *Hypothesis: transcript-templated repair of DNA double-strand breaks.* Bioessays, 2006. **28**(1): p. 78-83.

68. Diemer, G.S. and K.M. Stedman, *A novel virus genome discovered in an extreme environment suggests recombination between unrelated groups of RNA and DNA viruses.* Biol Direct, 2012. **7**: p. 13.

69. Krupovic, M., *Recombination between RNA viruses and plasmids might have played a central role in the origin and evolution of small DNA viruses.* Bioessays, 2012. **34**(10): p. 867-70.

70. Xu, P.Z., et al., *Genome-wide high-frequency non-Mendelian loss of heterozygosity in rice.* Genome, 2007. **50**(3): p. 297-302.

71. Suberbielle, E., et al., *Physiologic brain activity causes DNA double-strand breaks in neurons, with exacerbation by amyloid-beta.* Nat Neurosci, 2013. **16**(5): p. 613-21.

72. Mehler, M.F., *Epigenetics and the nervous system.* Ann Neurol, 2008. **64**(6): p. 602-17.

73. Mattick, J.S. and M.F. Mehler, *RNA editing, DNA recoding and the evolution of human cognition.* Trends Neurosci, 2008. **31**(5): p. 227-33.

74. Bozas, A., et al., *Genetic analysis of zinc-finger nuclease-induced gene targeting in Drosophila.* Genetics, 2009. **182**(3): p. 641-51.

75. Onozawa, M., et al., *Repair of DNA double-strand breaks by templated nucleotide sequence insertions derived from distant regions of the genome.* Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(21): p. 7729-7734.

76. Aida, T., et al., *Cloning-free CRISPR/Cas system facilitates functional cassette knock-in in mice.* Genome Biol, 2015. **16**: p. 87.

77. Shen, Y. and F. Storici, *Generation of RNA/DNA hybrids in genomic DNA by transformation using RNA-containing oligonucleotides.* J Vis Exp, 2010(45).

78. Shen, Y. and F. Storici, *Detection of RNA-Templated Double-Strand Break Repair in Yeast*, in *DNA Recombination*, H. Tsubouchi, Editor. 2011, Humana Press. p. 193-204.

79. Thaler, D.S., G. Tombline, and K. Zahn, *Short-patch reverse transcription in Escherichia coli.* Genetics, 1995. **140**(3): p. 909-15.

80. Shen, Y., et al., *RNA-driven genetic changes in bacteria and in human cells.* Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 2011. **717**(1–2): p. 91-98.

81. Angeleska, A., et al., *RNA-guided DNA assembly.* J Theor Biol, 2007. **248**(4): p. 706-20.

82. Kirkpatrick, D.P. and C.M. Radding, *RecA protein promotes rapid RNA-DNA hybridization in heterogeneous RNA mixtures.* Nucleic Acids Res, 1992. **20**(16): p. 4347-53.

83. Kirkpatrick, D.P., B.J. Rao, and C.M. Radding, *RNA-DNA hybridization promoted by E. coli RecA protein.* Nucleic Acids Res, 1992. **20**(16): p. 4339-46.

84. Kasahara, M., et al., *RecA protein-dependent R-loop formation in vitro.* Genes & Development, 2000. **14**(3): p. 360-365.

85. Zaitsev, E.N. and S.C. Kowalczykowski, *A novel pairing process promoted by Escherichia coli RecA protein: inverse DNA and RNA strand exchange.* Genes Dev, 2000. **14**(6): p. 740-9.

86. Wahba, L., S.K. Gore, and D. Koshland, *The homologous recombination machinery modulates the formation of RNA-DNA hybrids and associated chromosome instability.* Elife, 2013. **2**: p. e00505.

87. Hani, J. and H. Feldmann, *tRNA genes and retroelements in the yeast genome.* Nucleic Acids Res, 1998. **26**(3): p. 689-96.

88. Boeke, J.D., C.A. Styles, and G.R. Fink, *Saccharomyces cerevisiae SPT3 gene is required for transposition and transpositional recombination of chromosomal Ty elements.* Mol Cell Biol, 1986. **6**(11): p. 3575-81.

89. Cerritelli, S.M. and R.J. Crouch, *Ribonuclease H: the enzymes in eukaryotes.* Febs j, 2009. **276**(6): p. 1494-505.

90. Keskin, H. and F. Storici, *Defects in RNase H2 Stimulate DNA Break Repair by RNA Reverse Transcribed into cDNA.* Microrna, 2015. **4**(2): p. 109-16.

91. Symington, L.S., R. Rothstein, and M. Lisby, *Mechanisms and regulation of mitotic recombination in Saccharomyces cerevisiae.* Genetics, 2014. **198**(3): p. 795-835.

92. Kumar, A. and J.L. Bennetzen, *Plant retrotransposons.* Annu Rev Genet, 1999. **33**: p. 479-532.

93. Bannert, N. and R. Kurth, *Retroelements and the human genome: new perspectives on an old relation.* Proc Natl Acad Sci U S A, 2004. **101 Suppl 2**: p. 14572-9.

94. Heyer, W.-D., K.T. Ehmsen, and J. Liu, *Regulation of Homologous Recombination in Eukaryotes.* Annual Review of Genetics, 2010. **44**(1): p. 113-139.

95. Ceccaldi, R., B. Rondinelli, and A.D. D'Andrea, *Repair Pathway Choices and Consequences at the Double-Strand Break.* Trends in Cell Biology. **26**(1): p. 52-64.

96. Li, X. and W.D. Heyer, *Homologous recombination in DNA repair and DNA damage tolerance.* Cell Res, 2008. **18**(1): p. 99-113.

97. Crick, F., *On Protein Synthesis.* Symposia of the Society for Experimental Biology, 1958. **12**: p. 138-163.

98. Crick, F., *Central Dogma of Molecular Biology.* Nature, 1970. **227**(5258): p. 561-563.

99. Autexier, C. and N.F. Lue, *The Structure and Function of Telomerase Reverse Transcriptase.* Annual Review of Biochemistry, 2006. **75**(1): p. 493-517.

100. Esnault, C., J. Maestre, and T. Heidmann, *Human LINE retrotransposons generate processed pseudogenes.* Nature Genetics, 2000. **24**(4): p. 363-367.

101. Keskin, H., C. Meers, and F. Storici, *Transcript RNA supports precise repair of its own DNA gene.* RNA Biol, 2015: p. **In press**.

102. Morrish, T.A., et al., *DNA repair mediated by endonuclease-independent LINE-1 retrotransposition.* Nature Genetics, 2002. **31**(2): p. 159-165.

103. Friedl, A.A., et al., *Ty1 integrase overexpression leads to integration of non-Ty1 DNA fragments into the genome of Saccharomyces cerevisiae.* Mol Genet Genomics, 2010. **284**(4): p. 231-42.

104. Ding, W., et al., *L1 elements, processed pseudogenes and retrogenes in mammalian genomes.* IUBMB Life, 2006. **58**(12): p. 677-85.

105. Morrish, T.A., et al., *Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres.* Nature, 2007. **446**(7132): p. 208-212.

106. Sen, S.K., et al., *Endonuclease-independent insertion provides an alternative pathway for L1 retrotransposition in the human genome.* Nucleic Acids Research, 2007. **35**(11): p. 3741-3751.

107. Lesage, P. and A.L. Todeschini, *Happy together: the life and times of Ty retrotransposons and their hosts.* Cytogenet Genome Res, 2005. **110**(1-4): p. 70-90.

108. Nevo-Caspi, Y. and M. Kupiec, *Induction of Ty recombination in yeast by cDNA and transcription: role of the RAD1 and RAD52 genes.* Genetics, 1996. **144**(3): p. 947-55.

109. Parket, A., O. Inbar, and M. Kupiec, *Recombination of Ty elements in yeast can be induced by a double-strand break.* Genetics, 1995. **140**(1): p. 67-77.

110. Ma, W.P. and R.J. Crouch, *Escherichia coli RNase HI inhibits murine leukaemia virus reverse transcription in vitro and yeast retrotransposon Ty1 transposition in vivo.* Genes Cells, 1996. **1**(6): p. 581-93.

111. Lim, Y.W., et al., *Genome-wide DNA hypomethylation and RNA:DNA hybrid accumulation in Aicardi-Goutieres syndrome.* Elife, 2015. **4**.

112. El Hage, A., et al., *Genome-wide distribution of RNA-DNA hybrids identifies RNase H targets in tRNA genes, retrotransposons and mitochondria.* PLoS Genet, 2014. **10**(10): p. e1004716.

113. Downs, J.A. and S.P. Jackson, *Involvement of DNA end-binding protein Ku in Ty element retrotransposition.* Mol Cell Biol, 1999. **19**(9): p. 6260-8.

114. Chung, W.H., et al., *Defective resection at DNA double-strand breaks leads to de novo telomere formation and enhances gene targeting.* PLoS Genet, 2010. **6**(5): p. e1000948.

115. Shen, Y., et al., *RNA-driven genetic changes in bacteria and in human cells.* Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis, 2011. **717**(1-2): p. 91-98.

116. Kasahara, M., et al., *RecA protein-dependent R-loop formation in vitro.* Genes Dev, 2000. **14**(3): p. 360-5.

117. Aymard, F., et al., *Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks.* Nat Struct Mol Biol, 2014. **21**(4): p. 366-74.

118. Wei, L., et al., *DNA damage during the G0/G1 phase triggers RNA-templated, Cockayne syndrome B-dependent homologous recombination.* Proc Natl Acad Sci U S A, 2015. **112**(27): p. E3495-504.

119. Chaurasia, P., et al., *Preferential repair of DNA double-strand break at the active gene in vivo.* J Biol Chem, 2012. **287**(43): p. 36414-22.

120. Meers, C., H. Keskin, and F. Storici, *DNA repair by RNA: Templated, or not templated, that is the question.* DNA Repair (Amst), 2016. **44**: p. 17-21.

121. Jividen, K. and H. Li, *Chimeric RNAs generated by intergenic splicing in normal and cancer cells.* Genes Chromosomes Cancer, 2014. **53**(12): p. 963-71.

122. Li, H., et al., *A neoplastic gene fusion mimics trans-splicing of RNAs in normal human cells.* Science, 2008. **321**(5894): p. 1357-61.

123. Li, H., et al., *Gene fusions and RNA trans-splicing in normal and neoplastic human cells.* Cell Cycle, 2009. **8**(2): p. 218-22.

124. Kowarz, E., T. Dingermann, and R. Marschalek, *Do non-genomically encoded fusion transcripts cause recurrent chromosomal translocations?* Cancers (Basel), 2012. **4**(4): p. 1036-49.

125. Kowarz, E., et al., *Premature transcript termination, trans-splicing and DNA repair: a vicious path to cancer.* Am J Blood Res, 2011. **1**(1): p. 1-12.

126. Chakraborty, A., et al., *Classical non-homologous end-joining pathway utilizes nascent RNA for error-free double-strand break repair of transcribed genes.* Nat Commun, 2016. **7**: p. 13049.

127. Shi, C., et al., *Innate Reverse Transcriptase Activity of DNA Polymerase for Isothermal RNA Direct Detection.* J Am Chem Soc, 2015. **137**(43): p. 13804-6.

128. Keskin, H., C. Meers, and F. Storici, *Transcript RNA supports precise repair of its own DNA gene.* RNA Biol, 2016. **13**(2): p. 157-65.

129. Santos-Pereira, J.M. and A. Aguilera, *R loops: new modulators of genome dynamics and function.* Nat Rev Genet, 2015. **16**(10): p. 583-97.

130. Aguilera, A. and T. Garcia-Muse, *R loops: from transcription byproducts to threats to genome stability.* Mol Cell, 2012. **46**(2): p. 115-24.

131. Castellano-Pozo, M., et al., *R loops are linked to histone H3 S10 phosphorylation and chromatin condensation.* Mol Cell, 2013. **52**(4): p. 583-90.

132. Garcia-Pichardo, D., et al., *Histone Mutants Separate R Loop Formation from Genome Instability Induction.* Mol Cell, 2017. **66**(5): p. 597-609.e5.

133. Luger, K., et al., *Crystal structure of the nucleosome core particle at 2.8 A resolution.* Nature, 1997. **389**(6648): p. 251-60.

134. Soshnev, A.A., S.Z. Josefowicz, and C.D. Allis, *Greater Than the Sum of Parts: Complexity of the Dynamic Epigenome.* Mol Cell, 2016. **62**(5): p. 681-94.

135. Happel, N. and D. Doenecke, *Histone H1 and its isoforms: contribution to chromatin structure and function.* Gene, 2009. **431**(1-2): p. 1-12.

136. Bradbury, E.M., et al., *Studies on the role and mode of operation of the very-lysine-rich histone H1 (F1) in eukaryote chromatin. The conformation of histone H1.* Eur J Biochem, 1975. **52**(3): p. 605-13.

137. Chapman, G.E., P.G. Hartman, and E.M. Bradbury, *Studies on the role and mode of operation of the very-lysine-rich histone H1 in eukaryote chromatin. The isolation of the globular and non-globular regions of the histone H1 molecule.* Eur J Biochem, 1976. **61**(1): p. 69-75.

138. Allan, J., et al., *The structure of histone H1 and its location in chromatin.* Nature, 1980. **288**(5792): p. 675-9.

139. Ramakrishnan, V., et al., *Crystal structure of globular domain of histone H5 and its implications for nucleosome binding.* Nature, 1993. **362**(6417): p. 219-23.

140. Bassett, A., et al., *The folding and unfolding of eukaryotic chromatin.* Curr Opin Genet Dev, 2009. **19**(2): p. 159-65.

141. Gajiwala, K.S. and S.K. Burley, *Winged helix proteins.* Curr Opin Struct Biol, 2000. **10**(1): p. 110-6.

142. Escher, D. and W. Schaffner, *Gene activation at a distance and telomeric silencing are not affected by yeast histone H1.* Mol Gen Genet, 1997. **256**(4): p. 456-61.

143. Ono, K., et al., *The linker histone homolog Hho1p from Saccharomyces cerevisiae represents a winged helix-turn-helix fold as determined by NMR spectroscopy.* Nucleic Acids Res, 2003. **31**(24): p. 7199-207.

144. Schafer, G., E.M. Smith, and H.G. Patterton, *The Saccharomyces cerevisiae linker histone Hho1p, with two globular domains, can simultaneously bind to two four-way junction DNA molecules.* Biochemistry, 2005. **44**(50): p. 16766-75.

145. Varga-Weisz, P., K. van Holde, and J. Zlatanova, *Preferential binding of histone H1 to four-way helical junction DNA.* J Biol Chem, 1993. **268**(28): p. 20699-700.

146. Varga-Weisz, P., et al., *Binding of histones H1 and H5 and their globular domains to four-way junction DNA.* Proc Natl Acad Sci U S A, 1994. **91**(9): p. 3525-9.

147. Fan, Y., et al., *Individual somatic H1 subtypes are dispensable for mouse development even in mice lacking the H1(0) replacement subtype.* Mol Cell Biol, 2001. **21**(23): p. 7933-43.

148. Vujatovic, O., et al., *Drosophila melanogaster linker histone dH1 is required for transposon silencing and to preserve genome integrity.* Nucleic Acids Res, 2012. **40**(12): p. 5402-14.

149. Lu, X., et al., *Drosophila H1 regulates the genetic activity of heterochromatin by recruitment of Su(var)3-9.* Science, 2013. **340**(6128): p. 78-81.

150. Bayona-Feliu, A., et al., *Linker histone H1 prevents R-loop accumulation and genome instability in heterochromatin.* Nat Commun, 2017. **8**(1): p. 283.

151. Patterton, H.G., et al., *The biochemical and phenotypic characterization of Hho1p, the putative linker histone H1 of Saccharomyces cerevisiae.* J Biol Chem, 1998. **273**(13): p. 7268-76.

152. Freidkin, I. and D.J. Katcoff, *Specific distribution of the Saccharomyces cerevisiae linker histone homolog HHO1p in the chromatin.* Nucleic Acids Res, 2001. **29**(19): p. 4043-51.

153. Georgieva, M., et al., *Hho1p, the linker histone of Saccharomyces cerevisiae, is important for the proper chromatin organization in vivo.* Biochim Biophys Acta, 2012. **1819**(5): p. 366-74.

154. Schafer, G., C.R. McEvoy, and H.G. Patterton, *The Saccharomyces cerevisiae linker histone Hho1p is essential for chromatin compaction in stationary phase and is displaced by transcription.* Proc Natl Acad Sci U S A, 2008. **105**(39): p. 14838-43.

155. Li, C., et al., *Linker histone H1 represses recombination at the ribosomal DNA locus in the budding yeast Saccharomyces cerevisiae.* Mol Microbiol, 2008. **67**(4): p. 906-19.

156. Levy, A., et al., *Yeast linker histone Hho1p is required for efficient RNA polymerase I processivity and transcriptional silencing at the ribosomal DNA.* Proc Natl Acad Sci U S A, 2008. **105**(33): p. 11703-8.

157. Downs, J.A., et al., *Suppression of homologous recombination by the Saccharomyces cerevisiae linker histone.* Mol Cell, 2003. **11**(6): p. 1685-92.

158. Brush, G.S., *Evidence that histone H1 is dispensable for proper meiotic recombination in budding yeast.* BMC Res Notes, 2015. **8**: p. 275.

159. Bryant, J.M., et al., *The linker histone plays a dual role during gametogenesis in Saccharomyces cerevisiae.* Mol Cell Biol, 2012. **32**(14): p. 2771-83.

160. Dewhurst, H.M. and M.P. Torres, *Systematic analysis of non-structural protein features for the prediction of PTM function potential by artificial neural networks.* PLoS One, 2017. **12**(2): p. e0172572.

161. English, N. and M. Torres, *SAPH-ire TFx: A Machine Learning Recommendation Method and Webtool for the Prediction of Functional Post-Translational Modifications.* bioRxiv, 2019: p. 731026.

162. Torres, M.P., H. Dewhurst, and N. Sundararaman, *Proteome-wide Structural Analysis of PTM Hotspots Reveals Regulatory Elements Predicted to Impact Biological Function and Disease.* Mol Cell Proteomics, 2016. **15**(11): p. 3513-3528.

163. Bairoch, A., et al., *The Universal Protein Resource (UniProt).* Nucleic Acids Res, 2005. **33**(Database issue): p. D154-9.

164. Mitchell, A., et al., *The InterPro protein families database: the classification resource after 15 years.* Nucleic Acids Res, 2015. **43**(Database issue): p. D213-21.

165. Huang, K.Y., et al., *dbPTM 2016: 10-year anniversary of a resource for post-translational modification of proteins.* Nucleic Acids Res, 2016. **44**(D1): p. D435-46.

166. Hornbeck, P.V., et al., *PhosphoSitePlus, 2014: mutations, PTMs and recalibrations.* Nucleic Acids Res, 2015. **43**(Database issue): p. D512-20.

167. Edgar, R.C., *MUSCLE: a multiple sequence alignment method with reduced time and space complexity.* BMC Bioinformatics, 2004. **5**: p. 113.

168. Stuckey, S., K. Mukherjee, and F. Storici, *In vivo site-specific mutagenesis and gene collage using the delitto perfetto system in yeast Saccharomyces cerevisiae.* Methods Mol Biol, 2011. **745**: p. 173-91.

169. Storici, F., et al., *Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast.* Proc Natl Acad Sci U S A, 2003. **100**(25): p. 14994-9.

170. Wach, A., et al., *New heterologous modules for classical or PCR-based gene disruptions in Saccharomyces cerevisiae.* Yeast, 1994. **10**(13): p. 1793-808.

171. Ivanov, E.L., et al., *Mutations in XRS2 and RAD50 delay but do not prevent mating-type switching in Saccharomyces cerevisiae.* Mol Cell Biol, 1994. **14**(5): p. 3414-25.

172. Lee, M.J. and H.G. Dohlman, *Coactivation of G protein signaling by cell-surface receptors and an intracellular exchange factor.* Curr Biol, 2008. **18**(3): p. 211-5.

173. Schindelin, J., et al., *The ImageJ ecosystem: An open platform for biomedical image analysis.* Mol Reprod Dev, 2015. **82**(7-8): p. 518-29.

174. Shevchenko, A., et al., *In-gel digestion for mass spectrometric characterization of proteins and proteomes.* Nat Protoc, 2006. **1**(6): p. 2856-60.

175. Daujat, S., et al., *HP1 binds specifically to Lys26-methylated histone H1.4, whereas simultaneous Ser27 phosphorylation blocks HP1 binding.* J Biol Chem, 2005. **280**(45): p. 38090-5.

176. Liu, X., et al., *HIST1H1C Regulates Interferon-beta and Inhibits Influenza Virus Replication by Interacting with IRF3.* Front Immunol, 2017. **8**: p. 350.

177. Chu, C.S., et al., *Protein kinase A-mediated serine 35 phosphorylation dissociates histone H1.4 from mitotic chromosome.* J Biol Chem, 2011. **286**(41): p. 35843-51.

178. Roth, S.Y., et al., *Characterization of phosphorylation sites in histone H1 in the amitotic macronucleus of Tetrahymena during different physiological states.* J Cell Biol, 1988. **107**(6 Pt 2): p. 2473-82.

179. Dou, Y. and M.A. Gorovsky, *Phosphorylation of linker histone H1 regulates gene expression in vivo by creating a charge patch.* Mol Cell, 2000. **6**(2): p. 225-31.

180. Kim, K., et al., *Functional interplay between p53 acetylation and H1.2 phosphorylation in p53-regulated transcription.* Oncogene, 2012. **31**(39): p. 4290-301.

181. O'Brien, S.K., et al., *P-TEFb kinase complex phosphorylates histone H1 to regulate expression of cellular and HIV-1 genes.* J Biol Chem, 2010. **285**(39): p. 29713-20.

182. Albuquerque, C.P., et al., *A multidimensional chromatography technology for in-depth phosphoproteome analysis.* Mol Cell Proteomics, 2008. **7**(7): p. 1389-96.

183. Holt, L.J., et al., *Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution.* Science, 2009. **325**(5948): p. 1682-6.

184. Soulard, A., et al., *The rapamycin-sensitive phosphoproteome reveals that TOR controls protein kinase A toward some but not all substrates.* Mol Biol Cell, 2010. **21**(19): p. 3475-86.

185. Swaney, D.L., et al., *Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation.* Nat Methods, 2013. **10**(7): p. 676-82.

186. Haber, J.E., *A Life Investigating Pathways That Repair Broken Chromosomes.* Annu Rev Genet, 2016. **50**: p. 1-28.

187. Storici, F., et al., *Conservative repair of a chromosomal double-strand break by single-strand DNA through two steps of annealing.* Mol Cell Biol, 2006. **26**(20): p. 7645-57.

188. Sancar, A., et al., *Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints.* Annu Rev Biochem, 2004. **73**: p. 39-85.

189. Lundin, C., et al., *Methyl methanesulfonate (MMS) produces heat-labile DNA damage but no detectable in vivo DNA double-strand breaks.* Nucleic Acids Res, 2005. **33**(12): p. 3799-811.

190. Ma, W., M.A. Resnick, and D.A. Gordenin, *Apn1 and Apn2 endonucleases prevent accumulation of repair-associated DNA breaks in budding yeast as revealed by direct chromosomal analysis.* Nucleic Acids Res, 2008. **36**(6): p. 1836-46.

191. Ma, W., et al., *Alkylation base damage is converted into repairable double-strand breaks and complex intermediates in G2 cells lacking AP endonuclease.* PLoS Genet, 2011. **7**(4): p. e1002059.

192. Rastogi, R.P., et al., *Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair.* J Nucleic Acids, 2010. **2010**: p. 592980.

193. Song, X. and M.A. Gorovsky, *Unphosphorylated H1 is enriched in a specific region of the promoter when CDC2 is down-regulated during starvation.* Mol Cell Biol, 2007. **27**(5): p. 1925-33.

194. Clouaire, T., et al., *Comprehensive Mapping of Histone Modifications at DNA Double-Strand Breaks Deciphers Repair Pathway Chromatin Signatures.* Mol Cell, 2018. **72**(2): p. 250-262.e6.

195. Watts, F.Z., *Repair of DNA Double-Strand Breaks in Heterochromatin.* Biomolecules, 2016. **6**(4).

196. Downs, J.A., M.C. Nussenzweig, and A. Nussenzweig, *Chromatin dynamics and the preservation of genetic information.* Nature, 2007. **447**(7147): p. 951-8.

197. Bhargava, R., D.O. Onyango, and J.M. Stark, *Regulation of Single-Strand Annealing and its Role in Genome Maintenance.* Trends Genet, 2016. **32**(9): p. 566-575.

198. Bleykasten-Grosshans, C., A. Friedrich, and J. Schacherer, *Genome-wide analysis of intraspecific transposon diversity in yeast.* BMC Genomics, 2013. **14**: p. 399.

199. Chatterjee, N. and G.C. Walker, *Mechanisms of DNA damage, repair, and mutagenesis.* Environ Mol Mutagen, 2017. **58**(5): p. 235-263.

200. Jasin, M. and R. Rothstein, *Repair of strand breaks by homologous recombination.* Cold Spring Harb Perspect Biol, 2013. **5**(11): p. a012740.

201. Piazza, A. and W.D. Heyer, *Moving forward one step back at a time: reversibility during homologous recombination.* Curr Genet, 2019.

202. Goodier, J.L., *Restricting retrotransposons: a review.* Mobile DNA, 2016. **7**: p. 16-16.

203. DiCarlo, J.E., et al., *Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems.* Nucleic Acids Res, 2013. **41**(7): p. 4336-43.

204. Garfinkel, D.J., et al., *Ty1 copy number dynamics in Saccharomyces.* Genetics, 2005. **169**(4): p. 1845-1857.

205. Storici, F., L.K. Lewis, and M.A. Resnick, *In vivo site-directed mutagenesis using oligonucleotides.* Nat Biotechnol, 2001. **19**(8): p. 773-6.

206. Garfinkel, D.J., et al., *Transposon tagging using Ty elements in yeast.* Genetics, 1988. **120**(1): p. 95-108.

207. Sokal, R.R. and F.J. Rohlf, *Biometry : the principles and practice of statistics in biological research*. 2012, New York: W.H. Freeman.

208. Garfinkel, D.J., et al., *Proteolytic processing of pol-TYB proteins from the yeast retrotransposon Ty1*. Journal of virology, 1991. **65**(9): p. 4573-4581.

209. Xue, Y., et al., *A DNA microarray for fission yeast: minimal changes in global gene expression after temperature shift.* Yeast, 2004. **21**(1): p. 25-39.

210. Banyai, G., et al., *Cdk1 activity acts as a quantitative platform for coordinating cell cycle progression with periodic transcription.* Nature Communications, 2016. **7**(1): p. 11161.

211. Lyndaker, A.M. and E. Alani, *A tale of tails: insights into the coordination of 3' end processing during homologous recombination.* BioEssays : news and reviews in molecular, cellular and developmental biology, 2009. **31**(3): p. 315-321.

212. Ivanov, E.L. and J.E. Haber, *RAD1 and RAD10, but not other excision repair genes, are required for double-strand break-induced recombination in Saccharomyces cerevisiae.* Molecular and Cellular Biology, 1995. **15**(4): p. 2245-2251.

213. Sugawara, N., et al., *Role of Saccharomyces cerevisiae Msh2 and Msh3 repair proteins in double-strand break-induced recombination.* Proceedings of the National Academy of Sciences, 1997. **94**(17): p. 9214-9219.

214. Paques, F. and J.E. Haber, *Two pathways for removal of nonhomologous DNA ends during double-strand break repair in Saccharomyces cerevisiae.* Mol Cell Biol, 1997. **17**(11): p. 6765-71.

215. Sugawara, N., et al., *Heteroduplex rejection during single-strand annealing requires Sgs1 helicase and mismatch repair proteins Msh2 and Msh6 but not Pms1*. Proc Natl Acad Sci U S A, 2004. **101**(25): p. 9315-20.

216. McVey, M., et al., *Eukaryotic DNA Polymerases in Homologous Recombination.* Annual review of genetics, 2016. **50**: p. 393-421.

217. Kawasaki, Y. and A. Sugino, *Yeast replicative DNA polymerases and their role at the replication fork.* Mol Cells, 2001. **12**(3): p. 277-85.

218. Nelson, J.R., C.W. Lawrence, and D.C. Hinkle, *Deoxycytidyl transferase activity of yeast REV1 protein.* Nature, 1996. **382**(6593): p. 729-31.

219. Lazzaro, F., et al., *RNase H and postreplication repair protect cells from ribonucleotides incorporated in DNA.* Mol Cell, 2012. **45**(1): p. 99-110.

220. Su, Y., et al., *Human DNA polymerase eta has reverse transcriptase activity in cellular environments.* J Biol Chem, 2019. **294**(15): p. 6073-6081.

221. Makarova, A.V. and P.M. Burgers, *Eukaryotic DNA polymerase zeta.* DNA Repair (Amst), 2015. **29**: p. 47-55.

222. Gallo, D., et al., *Rad5 Recruits Error-Prone DNA Polymerases for Mutagenic Repair of ssDNA Gaps on Undamaged Templates.* Molecular cell, 2019. **73**(5): p. 900-914.e9.

223. Stone, J.E., et al., *Low-fidelity DNA synthesis by the L979F mutator derivative of Saccharomyces cerevisiae DNA polymerase ζ.* Nucleic Acids Research, 2009. **37**(11): p. 3774-3787.

224. Freudenreich, C.H., *R-loops: targets for nuclease cleavage and repeat instability.* Curr Genet, 2018. **64**(4): p. 789-794.

225. Sollier, J., et al., *Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability.* Mol Cell, 2014. **56**(6): p. 777-85.

226. Su, X.A. and C.H. Freudenreich, *Cytosine deamination and base excision repair cause R-loop-induced CAG repeat fragility and instability in Saccharomyces cerevisiae.* Proc Natl Acad Sci U S A, 2017. **114**(40): p. E8392-e8401.

227. Rondon, A.G. and A. Aguilera, *What causes an RNA-DNA hybrid to compromise genome integrity?* DNA Repair (Amst), 2019. **81**: p. 102660.

228. Yan, Z., et al., *Rad52 Restrains Resection at DNA Double-Strand Break Ends in Yeast.* Mol Cell, 2019.

229. Martin, S.K. and R.D. Wood, *DNA polymerase zeta in DNA replication and repair*. Nucleic Acids Res, 2019. **47**(16): p. 8348-8361.

230. Holmes, A.M. and J.E. Haber, *Double-strand break repair in yeast requires both leading and lagging strand DNA polymerases.* Cell, 1999. **96**(3): p. 415-24.

231. Maloisel, L., F. Fabre, and S. Gangloff, *DNA polymerase delta is preferentially recruited during homologous recombination to promote heteroduplex DNA extension.* Mol Cell Biol, 2008. **28**(4): p. 1373-82.

232. Li, X., et al., *PCNA is required for initiation of recombination-associated DNA synthesis by DNA polymerase delta.* Mol Cell, 2009. **36**(4): p. 704-13.

233. Sneeden, J.L., et al., *Reconstitution of recombination-associated DNA synthesis with human proteins.* Nucleic Acids Res, 2013. **41**(9): p. 4913-25.

234. Rattray, A.J., et al., *The roles of REV3 and RAD57 in double-strand-break-repair-induced mutagenesis of Saccharomyces cerevisiae*. Genetics, 2002. **162**(3): p. 1063-77.

235. Hirano, Y. and K. Sugimoto, *ATR homolog Mec1 controls association of DNA polymerase zeta-Rev1 complex with regions near a double-strand break.* Curr Biol, 2006. **16**(6): p. 586-90.

236. Zhong, X., et al., *The fidelity of DNA synthesis by yeast DNA polymerase zeta alone and with accessory proteins.* Nucleic acids research, 2006. **34**(17): p. 4731-4742.

237. Wimberly, H., et al., *R-loops and nicks initiate DNA breakage and genome instability in non-growing Escherichia coli.* Nat Commun, 2013. **4**: p. 2115.

238. Stuckey, R., et al., *Role for RNA:DNA hybrids in origin-independent replication priming in a eukaryotic system.* Proceedings of the National Academy of Sciences, 2015. **112**(18): p. 5779-5784.

239. Posse, V., et al., *RNase H1 directs origin-specific initiation of DNA replication in human mitochondria.* PLoS Genet, 2019. **15**(1): p. e1007781.

240. Rodriguez, G.P., et al., *Mismatch repair-dependent mutagenesis in nondividing cells.* Proc Natl Acad Sci U S A, 2012. **109**(16): p. 6153-8.

241. Feschotte, C. and E.J. Pritham, *DNA transposons and the evolution of eukaryotic genomes.* Annu Rev Genet, 2007. **41**: p. 331-68.

242. Ono, R., et al., *Double strand break repair by capture of retrotransposon sequences and reverse-transcribed spliced mRNA sequences in mouse zygotes.* Sci Rep, 2015. **5**: p. 12281.

243. Onozawa, M., et al., *Repair of DNA double-strand breaks by templated nucleotide sequence insertions derived from distant regions of the genome.* Proc Natl Acad Sci U S A, 2014. **111**(21): p. 7729-34.

244. Lee, M.H., et al., *Somatic APP gene recombination in Alzheimer's disease and normal neurons.* Nature, 2018. **563**(7733): p. 639-645.

245. Nevo-Caspi, Y. and M. Kupiec, *Induction of Ty recombination in yeast by cDNA and transcription: role of the RAD1 and RAD52 genes.* Genetics, 1996. **144**(3): p. 947-955.

246. Mules, E.H., O. Uzun, and A. Gabriel, *In vivo Ty1 reverse transcription can generate replication intermediates with untidy ends.* Journal of virology, 1998. **72**(8): p. 6490-6503.

247. Dombroski, B.A., et al., *An in vivo assay for the reverse transcriptase of human retrotransposon L1 in Saccharomyces cerevisiae.* Molecular and Cellular Biology, 1994. **14**(7): p. 4485.

248. Warren, I.A., et al., *Evolutionary impact of transposable elements on genomic diversity and lineage-specific innovation in vertebrates.* Chromosome Res, 2015. **23**(3): p. 505-31.

249. Yamashiro, H. and M.C. Siomi, *PIWI-Interacting RNA in Drosophila: Biogenesis, Transposon Regulation, and Beyond.* Chem Rev, 2018. **118**(8): p. 4404-4421.

250. Hedges, D.J. and P.L. Deininger, *Inviting instability: Transposable elements, double-strand breaks, and the maintenance of genome integrity.* Mutation research, 2007. **616**(1-2): p. 46-59.

251. Steele, E.J., et al., *Genesis of ancestral haplotypes: RNA modifications and reverse transcription–mediated polymorphisms.* Human immunology, 2011. **72**(3): p. 283-293. e1.

252. Chen, X. and L.F. Landweber, *Phylogenomic analysis reveals genome-wide purifying selection on TBE transposons in the ciliate Oxytricha.* Mobile DNA, 2016. **7**(1): p. 2.

253. Koonin, E.V. and K.S. Makarova, *Mobile Genetic Elements and Evolution of CRISPR-Cas Systems: All the Way There and Back.* Genome Biol Evol, 2017. **9**(10): p. 2812-2825.

254. Jangam, D., C. Feschotte, and E. Betrán, *Transposable Element Domestication As an Adaptation to Evolutionary Conflicts.* Trends in genetics : TIG, 2017. **33**(11): p. 817-831.

255. Zhang, Y., et al., *Transposon molecular domestication and the evolution of the RAG recombinase.* Nature, 2019. **569**(7754): p. 79-84.

256. Friant, S., et al., *Interactions between Ty1 retrotransposon RNA and the T and D regions of the tRNA(iMet) primer are required for initiation of reverse transcription in vivo.* Molecular and cellular biology, 1998. **18**(2): p. 799-806.

257. Farzadfard, F. and T.K. Lu, *Synthetic biology. Genomically encoded analog memory with precise in vivo DNA writing in living cell populations.* Science (New York, N.Y.), 2014. **346**(6211): p. 1256272-1256272.

258. Beck, C.R., et al., *LINE-1 retrotransposition activity in human genomes.* Cell, 2010. **141**(7): p. 1159-1170.

259. Mason, J.M., R.C. Frydrychova, and H. Biessmann, *Drosophila telomeres: an exception providing new insights.* BioEssays : news and reviews in molecular, cellular and developmental biology, 2008. **30**(1): p. 25-37.